

(19)



(11)

EP 2 525 355 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:
01.11.2017 Bulletin 2017/44

(51) Int Cl.:
G10L 19/02 ^(2013.01) **G10L 19/24** ^(2013.01)
G10L 19/12 ^(2013.01)

(21) Application number: **11732797.3**

(86) International application number:
PCT/JP2011/000134

(22) Date of filing: **13.01.2011**

(87) International publication number:
WO 2011/086924 (21.07.2011 Gazette 2011/29)

(54) AUDIO ENCODING APPARATUS AND AUDIO ENCODING METHOD

AUDIOKODIERUNGSVORRICHTUNG UND AUDIOKODIERUNGSVERFAHREN

APPAREIL DE CODAGE AUDIO ET PROCÉDÉ DE CODAGE AUDIO

(84) Designated Contracting States:
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO
PL PT RO RS SE SI SK SM TR**

(30) Priority: **14.01.2010 JP 2010006312**

(43) Date of publication of application:
21.11.2012 Bulletin 2012/47

(73) Proprietor: **Panasonic Intellectual Property
Corporation
of America
Torrance, CA 90503 (US)**

(72) Inventor: **LIU, Zongxian
Osaka-shi, Osaka 540 - 6207 (JP)**

(74) Representative: **Grünecker Patent- und
Rechtsanwälte
PartG mbB
Leopoldstraße 4
80802 München (DE)**

(56) References cited:
**WO-A1-2006/075563 WO-A1-2009/039645
JP-A- 10 020 898 JP-A- 2001 331 198
JP-A- 2006 011 170 JP-A- 2007 193 043
US-A1- 2007 016 404 US-A1- 2009 083 040
US-A1- 2009 240 491**

EP 2 525 355 B1

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

Technical Field

5 **[0001]** The present invention relates to an encoding speech apparatus and an encoding speech method.

Background Art

10 **[0002]** Speech coding techniques are categorized into mainly two coding techniques, i.e., transform coding and linear predictive coding.

15 **[0003]** The transform coding transforms signals from a time domain into a spectral domain and then encodes spectral coefficients using a discrete Fourier transform (DFT) or a modified discrete cosine transform (MDCT), for example. The coding process generally involves calculating perceptual importance levels of the spectral coefficients using a psycho-acoustic model and then encoding the spectral coefficients according to each perceptual importance level. Some common transform coding techniques include MPEG MP3, MPEG AAC, and Dolby AC3. The transform coding is effective for music signals and general speech signals.

20 **[0004]** FIG.1 illustrates a configuration of the transform coding.

25 **[0005]** In the coding side of FIG.1, time-frequency transforming section 101 transforms time domain signal $S(n)$ into frequency domain signal $S(f)$ using time-frequency transformation such as discrete Fourier transform (DFT) or modified discrete cosine transform (MDCT).

30 **[0006]** Psychoacoustic model analyzing section 103 performs a psychoacoustic model analysis on frequency domain signal $S(f)$ to calculate a masking curve.

35 **[0007]** According to the masking curve calculated by the psychoacoustic model analysis, coding section 102 encodes frequency domain signal $S(f)$ not to create quantization noise.

40 **[0008]** Multiplexing section 104 multiplexes the coding parameter generated at coding section 102 with the signal to generate bit stream information, and transmits the bit stream information to a decoding side.

45 **[0009]** In the decoding side of FIG.1, demultiplexing section 105 demultiplexes the bit stream information to generate the coding parameter.

50 **[0010]** Decoding section 106 decodes the coding parameter to generate decoded frequency domain signal $\hat{S}(f)$.

55 **[0011]** By using frequency-time transformation such as inverse discrete Fourier transform (IDFT) or inverse modified discrete cosine transform (IMDCT), frequency-time transforming section 107 transforms decoded frequency domain signal $\hat{S}(f)$ into a time domain, to generate decoded time domain signal $\hat{S}(n)$.

60 **[0012]** On the other hand, the linear predictive coding obtains a residual/excitation signal by using redundancy of a speech signal in a time domain and applying linear prediction to an input speech signal. In the case of a speech signal, especially an active speech section (with resonance effect and a component of a pitch period with high amplitude), the linear predictive coding efficiently generates an audio playback signal. After the linear prediction, main two different techniques such as TCX and CELP encode the residual/excitation signal.

65 **[0013]** TCX efficiently transforms and encodes the residual/excitation signal in a frequency domain. Some common TCX coding techniques include 3GPP AMR-WB+, MPEG USAC, for example.

70 **[0014]** FIG.2 illustrates a configuration of the TCX coding.

75 **[0015]** In the coding side of FIG.2, LPC analyzing section 201 performs LPC analysis on an input signal to use redundancy of a signal in a time domain.

80 **[0016]** Coding section 202 encodes the LPC coefficients from LPC analyzing section 201.

85 **[0017]** Decoding section 203 decodes the encoded LPC coefficients.

90 **[0018]** Inverse filter section 204 applies an LPC inverse filter to input signal $S(n)$, using the decoded LPC coefficients from decoding section 203, to obtain residual (excitation) signal $S_r(n)$.

95 **[0019]** Time-frequency transforming section 205 transforms residual signal $S_r(n)$ into frequency domain signal $S_r(f)$ using time-frequency transformation such as discrete Fourier transform (DFT) or modified discrete cosine transform (MDCT).

100 **[0020]** Coding section 206 encodes $S_r(f)$.

105 **[0021]** Multiplexing section 207 multiplexes the LPC coefficients generated and encoded at coding section 202 and the coding parameter generated at coding section 206 to generate bit stream information, and transmits the bit stream information to the decoding side.

110 **[0022]** In the decoding side of FIG.2, demultiplexing section 208 demultiplexes the bit stream information to generate the encoded LPC coefficients and coding parameter.

115 **[0023]** Decoding section 210 decodes the coding parameter to generate decoded residual signal $\hat{S}_r(f)$ of a frequency domain.

120 **[0024]** LPC coefficient decoding section 209 decodes the encoded LPC coefficients to obtain LPC coefficients.

[0025] By using frequency-time transformation such as inverse discrete Fourier transform (IDFT) or inverse modified discrete cosine transform (IMDCT), frequency-time transforming section 211 transforms decoded residual signal $S_r \sim (f)$ of a frequency domain into a time domain, to generate decoded residual signal $S_r \sim (n)$ of the time domain.

[0026] Synthesis filter 212 performs LPC synthesis filtering processing on decoded residual signal $S_r \sim (n)$ of the time domain using the LPC coefficients decoded at LPC coefficient decoding section 209, to obtain decoded time domain signal $S \sim (n)$.

[0027] Also, CELP coding encodes a residual/excitation signal using a predetermined code book. In order to improve the sound quality, the CELP coding transforms an error signal into a frequency domain for coding, the error signal between the original signal and an LPC synthesized signal. Common CELP coding techniques include ITU-T G.729.1, ITU-T G.718, for example.

[0028] FIG.3 illustrates a configuration of coding combining the CELP coding and the transform coding.

[0029] In the coding side of FIG.3, CELP coding section 301 performs the CELP coding on an input signal to use redundancy of a signal in a time domain.

[0030] CELP decoding section 302 generates synthesized signal $S_{syn}(n)$ using a CELP parameter generated at CELP coding section 301.

[0031] By subtracting the synthesized signal from the input signal, subtractor 310 obtains error signal $S_e(n)$ (error signal between the input signal and the synthesized signal).

[0032] Time-frequency transforming section 303 transforms error signal $S_e(n)$ into frequency domain signal $S_e(f)$ (spectral coefficients) using time-frequency transformation such as discrete Fourier transform (DFT) or modified discrete cosine transform (MDCT).

[0033] Coding section 304 encodes $S_e(f)$.

[0034] Multiplexing section 305 multiplexes the CELP parameter generated at CELP coding section 301 and the coding parameter generated at coding section 304 to generate bit stream information, and transmits the bit stream information to the decoding side.

[0035] In the decoding side of FIG.3, demultiplexing section 306 demultiplexes the bit stream information to generate the CELP parameter and the coding parameter.

[0036] Decoding section 308 decodes the coding parameter to generate decoded residual signal $S_e \sim (f)$ of a frequency domain.

[0037] CELP decoding section 307 generates CELP synthesized signal $S_{syn}(n)$ using the CELP parameter.

[0038] Frequency-time transforming section 309 transforms decoded residual signal $S_e \sim (f)$ of a frequency domain into a time domain using frequency-time transformation such as inverse discrete Fourier transform (IDFT) or inverse modified discrete cosine transform (IMDCT), to generate decoded residual signal (predictive error signal) $S_e \sim (n)$ of the time domain.

[0039] Adder 311 generates decoded time domain signal $S \sim (n)$ by adding CELP synthesized signal $S_{syn}(n)$ and decoded predictive error signal $S_e \sim (n)$.

[0040] Transform coding and linear predictive coding apply a certain coding technique to a signal of a frequency domain, that is, spectral coefficients (transform coefficients).

[0041] In order to concentrate limited coding bits to perceptually-important spectral coefficients, generally before encoding, coding of spectral coefficients by transform coding calculates weighting coefficients representing the perceptual importance level of the spectral coefficients, to use for encoding the spectral coefficients.

[0042] The transform coding generally calculates perceptually-weighting coefficients according to a psychoacoustic model to use masking phenomenon which is specific to human hearing mechanism.

[0043] Meanwhile, since the linear predictive coding performs linear prediction on an input signal, it is not easy to obtain a psychoacoustic model. Thus, the perceptually-weighting coefficients are generally calculated based on an energy-to-noise ratio or a signal-to-noise ratio.

[0044] Hereinafter, the coding of the spectral coefficients applied to the transform coding or the linear predictive coding is referred to as "pulse vector coding."

[0045] In the fifth layer in ITU-T G.718 which is newly-standardized speech coding, factorial pulse coding which is one of pulse vector coding technique has been proposed (FIG.4).

[0046] The factorial pulse coding is pulse vector coding in which coding information is a unit magnitude pulse. In the pulse vector coding, the spectral coefficients which are coding targets are represented by a plurality of pulses, and the positions, amplitudes, and polarities of these pulses are calculated, to encode this information. In this case, in order to normalize a pulse by unit amplitude, a global gain is also calculated for coding. As illustrated in FIG.5, the coding parameter of the pulse vector coding includes a global gain, a pulse position, a pulse amplitude, and a pulse polarity.

[0047] FIG.6 shows a concept of the pulse vector coding.

[0048] As illustrated in FIG.6, in input spectrum $S(f)$ having a length equal to N , one global gain and the positions, amplitudes, and polarities of M pulses, are encoded together. In spectrum $S \sim (f)$ generated by encoding, only M pulses and their positions, amplitudes, and polarities are generated and all the other spectral coefficients are set as zero.

[0049] Conventional transform coding calculates the perceptual importance level based on a subband. One example

is TDAC (Time Domain Aliasing Cancellation) coding in G.729.1.

[0050] FIG.7 illustrates a configuration of the TDAC coding in G.729.1.

[0051] In FIG.7, band splitting section 701 splits input signal $S(f)$ (spectral coefficients) into a plurality of subbands. Here, the low band section of the input signal is formed by error-signal MDCT coefficients between the original signal and a CELP decoded signal, and the high band section of the input signal is formed by MDCT coefficients of the original signal.

[0052] Spectrum envelope calculating section 702 calculates a spectrum envelope (energy of each subband) for each subband signal $\{S_{sb}(f)\}$.

[0053] Coding section 703 encodes the spectrum envelope.

[0054] Bit allocating section 704 calculates the order of perceptual importance levels $\{ip_{sb}\}$ according to the encoded spectrum envelopes, to allocate bits to subbands.

[0055] Vector quantizing section 705 uses the allocated bits and split spherical VQ method to encode subband signal $\{S_{sb}(f)\}$.

Citation List

Non-Patent Literature

[0056]

NPL 1 ITU-T Recommendation G.729.1 (2007) "G.729-based embedded variable bit-rate coder: An 8-32kbit/s scalable wideband coder bit stream interoperable with G.729"

NPL 2 T. Vaillancourt et al "ITU-T EV-VBR: A Robust 8-32 kbit/s Scalable Coder for Error Prone Telecommunication Channels," in Proc. Eusipco, Lausanne, Switzerland, August 2008

NPL 3 Lefebvre, et al., "High quality coding of wideband audio signals using transform coded excitation (TCX)," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. I/193-I/196, Apr. 1994

NPL 4 Karl Heinz Brandenburg, "MP3 and AAC Explained," AES 17th International Conference, Florence, Italy, September 1999.

Summary of Invention

Technical Problem

[0057] Here, it is not effective to calculate the perceptual importance level on a subband basis in a specific coding method such as the above mentioned pulse vector coding.

[0058] Calculating the perceptual importance level on a subband basis means that the all perceptual importance levels of the spectral coefficients included in each of the subbands are the same.

[0059] Meanwhile, from spectra of all bandwidth, pulse vector coding selects spectral coefficients to be encoded, based on amplitude values of spectral coefficients. In this case, the perceptual importance level calculated on a subband basis cannot accurately represent the perceptual importance level of spectral coefficients.

[0060] Let us consider a case where a certain subband includes five spectral coefficients $S_{sb}(f_0)$, $S_{sb}(f_1)$, $S_{sb}(f_2)$, $S_{sb}(f_3)$, and $S_{sb}(f_4)$ as illustrated in FIG.8. Also, pulse vector coding is adopted as a coding method in this case. Assuming that $S_{sb}(f_1)$ has the largest amplitude among the five spectral coefficients and coding bits allocated to this subband can encode only one pulse in this case, the pulse vector coding selects and encodes $S_{sb}(f_1)$. Here, even if the perceptual importance levels are calculated in this subband, $S_{sb}(f_1)$ is still encoded. This is because all the perceptual importance levels of five spectral coefficients are the same. However, calculating masking curve $M(f)$ of the original signal shows that $S_{sb}(f_3)$ exceeds masking curve $M(f)$, and hence it is understood that $S_{sb}(f_3)$ is the most perceptually-important spectral coefficient. Thus, when calculating the perceptual importance levels on a subband basis, a different spectral coefficient (in this example, $S_{sb}(f_1)$) with the largest amplitude value, is encoded, instead of encoding the most perceptually-important spectral coefficient (in this example, $S_{sb}(f_3)$).

[0061] Here, although there is a conventional technique determining the masking curve on a frequency basis, the technique performs the distribution of coding bits and perceptual weighting processing on a subband basis. That is, the differences among the perceptual importance levels of spectral coefficients included in a subband are not taken into consideration.

[0062] US 2007/016404 A1 discloses an method and apparatus to extract an audio signal having an important spectral component (ISC) and a low bit-rate audio signal coding/decoding method using the method and apparatus to extract the ISC. The method of extracting the ISC includes calculating perceptual importance including an SMR (signal-to-mask ratio) value of transformed spectral audio signals by using a psychoacoustic model, selecting spectral signals having a

masking threshold value smaller than that of the spectral audio signals using the SMR value as first ISCs, and extracting a spectral peak from the audio signals selected as the ISCs according to a predetermined weighting factor to select second ISCs.

5 Solution to Problem

[0063] The speech coding apparatus of the present invention employs a configuration having the features of claim 1 including at least two layers of a lower layer and a higher layer, employs a configuration having: a generating section that generates an error signal between a decoded signal of the lower layer and an input signal; an estimation section
10 that calculates a signal-to-noise ratio using the input signal and the error signal and estimates respective perceptual importance levels of a plurality of spectral coefficients of different frequencies in the error signal, based on the signal-to-noise ratio; a calculating section that calculates respective weighting coefficients of a plurality of spectral coefficients based on the respective estimated importance levels; a weighting section that weights each of a plurality of spectral coefficients using the respective calculated weighting coefficients; and a coding section that encodes a plurality of
15 weighted spectral coefficients.

[0064] The speech coding method of the present invention comprises the steps of claim 2.

Advantageous Effects of Invention

[0065] According to the present invention, the decoding side can obtain a decoded signal with good sound quality.

Brief Description of Drawings

[0066]

25 FIG.1 illustrates a configuration of transform coding (conventional);
FIG.2 illustrates a configuration of TCX coding (conventional);
FIG.3 illustrates a configuration of coding combining CELP coding and transform coding (conventional);
FIG.4 illustrates a configuration of factorial pulse coding of ITU-T G.718 (conventional);
30 FIG.5 illustrates a coding parameter of pulse vector coding (conventional);
FIG.6 illustrates a concept of the pulse vector coding (conventional);
FIG.7 illustrates a configuration of TDAC coding in G.729.1 (conventional);
FIG.8 illustrates a calculation example of perceptual importance level of the TDAC coding in G.729.1;
FIG.9 illustrates a calculation example of perceptual importance level of the present invention;
35 FIG.10A illustrates a configuration of a speech coding apparatus;
FIG.10B illustrates a speech decoding apparatus;
FIG.11 illustrates a configuration of a perceptually-weighting section;
FIG.12 illustrates a state where each spectral coefficient is perceptually weighted;
FIG.13A illustrates a configuration of a speech coding apparatus;
40 FIG. 13B illustrates a speech decoding apparatus;
FIG.14 illustrates a configuration of a perceptually-weighting section;
FIG.15 illustrates a state where each spectral coefficient is perceptually weighted;
FIG.16A illustrates a configuration of a speech coding apparatus according to an embodiment of the present invention;
FIG.16B illustrates a speech decoding apparatus;
45 FIG.17 illustrates a configuration of a perceptually-weighting section according to an example (configuration example 1);
FIG.18 illustrates a configuration of a perceptually-weighting section according to an embodiment of the present invention (configuration example 2); and
FIG.19 illustrates a state where each spectral coefficient is perceptually weighted in an embodiment of the present
50 invention.

Description of Embodiment

[0067] The present invention calculates the perceptual importance level, not on a subband basis but on each spectral
55 coefficient basis in encoding spectral coefficients. The present invention calculates respective weighting coefficients for applying the weighting coefficients to the spectral coefficients, according to a psychoacoustic model analysis, a signal-to-noise ratio, or the resulting perceptual importance levels based on a parameter related to a perceptual system. The weighting coefficient is larger as the perceptual importance level of a spectral coefficient is higher, and the weighting

coefficient is smaller as the perceptual importance level is lower. Thus, it is possible to obtain perceptually good sound quality by encoding a perceptually-weighted spectral coefficient.

[0068] The present invention determines the perceptual importance level according to a masking curve as illustrated in FIG.9. The perceptual importance level shows that $S_{sb}(f1)$ has the largest amplitude but is not perceptually important. For this reason, assignment of a low weight to $S_{sb}(f1)$ with low perceptual importance level suppresses $S_{sb}(f1)$. As a result, the most perceptually-important $S_{sb}(f3)$ will be encoded.

[0069] A first example determines respective perceptual importance levels of spectral coefficients, then determines weighting coefficients according to the perceptual importance levels, applies the weighting coefficients to the spectral coefficients, respectively, and encodes the perceptually-weighted spectral coefficients.

[0070] By this means, the perceptually-weighting coefficients are more accurate because the coefficients are calculated respectively for the spectral coefficients. It is therefore possible to select and encode the most perceptually-important spectral coefficient, and thereby to obtain better coding performance (improvement in sound quality).

[0071] In a second example, only the coding side applies the perceptually-weighting coefficients. That is, the decoding side does not perform inverse weighting processing corresponding to the application at the coding side.

[0072] By this means, there is no need to transmit the perceptually-weighting coefficients to the decoding side. Thus, it is possible to save bits in encoding the perceptually-weighting coefficients.

[0073] According to the present invention as claimed, layer coding (scalable coding) updates the perceptual importance level of an error signal in each layer. In each layer, the layer coding calculates the weight according to the perceptual importance level and applied to each coding-target spectral coefficient.

[0074] By this means, in each coding step or layer, a signal is encoded according to the perceptual importance level, and therefore it is possible to obtain better coding performance (improvement in sound quality) in each coding step or layer.

[0075] Examples and an embodiment of the present invention will now be explained with reference to the accompanying drawings.

(Example 1)

[0076] FIG.10A illustrates a configuration of speech coding apparatus 1000A according to an example. FIG.10B illustrates a configuration of speech decoding apparatus 1000B according to another example.

[0077] In the present example, a pulse vector coding perceptually weights each spectral coefficient.

[0078] In speech coding apparatus 1000A (FIG.10A), time-frequency transforming section 1001 transforms time domain signal $S(n)$ into frequency domain signal $S(f)$ (spectral coefficients), using time-frequency transformation such as discrete Fourier transform (DFT) or modified discrete cosine transform (MDCT).

[0079] Psychoacoustic model analyzing section 1002 determines a masking curve by performing a psychoacoustic model analysis on frequency domain signal $S(f)$.

[0080] Perceptually-weighting section 1003 estimates perceptual importance levels based on the masking curve, and calculates respective weighting coefficients for the spectral coefficients according to the perceptual importance levels, to apply the weighting coefficients to the spectral coefficients, respectively.

[0081] Coding section 1004 encodes perceptually-weighted frequency domain signal $S_{PW}(f)$ to generate a coding parameter.

[0082] Multiplexing section 1005 multiplexes the coding parameter with the signal to generate bit stream information and transmits the bit stream information to speech decoding apparatus 1000B (FIG.10B).

[0083] In speech decoding apparatus 1000B (FIG.10B), demultiplexing section 1006 demultiplexes the bit stream information to generate the coding parameter.

[0084] Decoding section 1007 decodes the coding parameter to generate decoded frequency domain signal $\hat{S}(f)$.

[0085] Frequency-time transforming section 1008 transforms decoded frequency domain signal $\hat{S}(f)$ into a time domain using frequency-time transformation such as inverse discrete Fourier transform (IDFT) or inverse modified discrete cosine transform (IMDCT), to generate decoded time domain signal $\hat{S}(n)$.

[0086] FIG.11 illustrates a configuration of perceptually-weighting section 1003 according to the present example. FIG.11 illustrates a configuration to perceptually weight each spectral coefficient.

[0087] In perceptually-weighting section 1003, estimation section 1101 estimates perceptual importance level $pi(f)$ of each spectral coefficient, according to masking curve $M(f)$. Perceptual importance level $pi(f)$ is the parameter quantitatively representing how perceptually important the spectral coefficient is. Perceptual importance level $pi(f)$ showing a larger value means that the spectral coefficient corresponding to the $pi(f)$ is perceptually important. Perceptual importance level $pi(f)$ is calculated based on masking curve $M(f)$ and an energy level of a spectral coefficient. The calculation may be performed in a logarithmic region, and, for example, perceptual importance level $pi(f)$ is calculated according to the following equation.

$$pi(f) = \log(S^2(f)) - \log(M(f)) \quad \dots \text{ (Equation 1)}$$

[0088] Weighting coefficient calculating section 1102 calculates weighting coefficient $W(f)$ based on perceptual importance level $pi(f)$. Weighting coefficient $W(f)$ is used for applying a weight to spectral coefficient $S(f)$. As perceptual importance level $pi(f)$ shows a larger value, weighting coefficient $W(f)$ becomes a larger value. For example, weighting coefficient $W(f)$ is calculated as the following equation.

$$W(f) = e^{pi(f)} \quad \dots \text{ (Equation 2)}$$

[0089] Weighting section 1103 multiplies spectral coefficient $S(f)$ by weighting coefficient $W(f)$ to generate perceptually-weighted spectral coefficient $S_{PW}(f)$. Thus, spectral coefficient $S_{PW}(f)$ is calculated as the following equation.

$$S_{PW}(f) = W(f) * S(f) \quad \dots \text{ (Equation 3)}$$

[0090] FIG.12 illustrates a state where each spectral coefficient is perceptually weighted.

[0091] As illustrated in FIG.12, energy levels of spectral coefficient $S(f0)$ and $S(f4)$ are lower than points of masking curve $M(f0)$ and $M(f1)$, respectively. At this time, weighting coefficients $W(f0)$ and $W(f4)$ multiplied to these two spectral coefficients respectively are less than 1, and hence the energy levels of spectral coefficients $S(f0)$ and $S(f4)$ are suppressed.

[0092] As an example, when perceptual importance level $pi(f)$ and weighting coefficient $W(f)$ are calculated as the above, perceptually-weighted spectral coefficients $S_{PW}(f0)$ and $S_{PW}(f4)$ are represented as the following, and reveal that those coefficients $S_{PW}(f0)$ and $S_{PW}(f4)$ become lower than spectral coefficients $S(f0)$ and $S(f4)$ respectively.

$$\begin{aligned} pi(f0) &= \log(S^2(f0)) - \log(M(f0)) < 0; \\ pi(f4) &= \log(S^2(f4)) - \log(M(f4)) < 0; \\ \text{then} \\ W(f0) &= e^{pi(f0)} < 1; \\ W(f4) &= e^{pi(f4)} < 1; \\ \text{then} \\ S_{PW}(f0) &= W(f0) * S(f0) < S(f0); \\ S_{PW}(f4) &= W(f4) * S(f4) < S(f4); \quad \dots \text{ (Equation 4)} \end{aligned}$$

[0093] According to the present example, a pulse vector coding determines the perceptual importance levels of the spectral coefficients, determines weighting coefficients according to the perceptual importance levels, applies the weighting coefficients to the respective spectral coefficients, and encodes the perceptually-weighted spectral coefficients.

[0094] By this means, the perceptually-weighting coefficients can calculate each spectral coefficient more accurately, in comparison with performing perceptual-weighting processing on a subband basis. Thus, it is possible to select and encode the most perceptually-important spectral coefficients and hence to obtain better coding performance.

[0095] According to the present example, only the coding side (speech coding apparatus 1000A) applies perceptually-weighting coefficients. That is, the decoding side (speech decoding apparatus 1000B) does not perform inverse weighting processing with respect to the above processing.

[0096] By this means, there is no need to transmit perceptually-weighting coefficients to the decoding side. Thus, it is possible to save bits in encoding the perceptually-weighting coefficients.

(Example 2)

[0097] FIG.13A illustrates a configuration of speech coding apparatus 1300A according to the present example.

FIG.13B also illustrates a configuration of speech decoding apparatus 1300B according to the present example.

[0098] In the present example, a TCX coding perceptually weights each spectral coefficient.

[0099] In speech coding apparatus 1300A (FIG.13A), LPC analyzing section 1301 performs LPC analysis on an input signal, so as to use redundancy of a signal in a time domain.

[0100] Coding section 1302 encodes the LPC coefficients from LPC analyzing section 1301.

[0101] Decoding section 1303 decodes the encoded LPC coefficients.

[0102] Inverse filter section 1304 obtains residual (excitation) signal $S_r(n)$ by applying an LPC inverse filter to input signal $S(n)$ using the decoded LPC coefficients from decoding section 1303.

[0103] Time-frequency transforming section 1305 transforms residual signal $S_r(n)$ into frequency domain signal $S_r(f)$ (spectral coefficients) using time-frequency transformation such as discrete Fourier transform (DFT) or modified discrete cosine transform (MDCT).

[0104] Time-frequency transforming section 1306 transforms original signal $S(n)$ into frequency domain signal $S(f)$ (spectral coefficients) using time-frequency transformation such as discrete Fourier transform (DFT) or modified discrete cosine transform (MDCT).

[0105] Perceptually-weighting section 1307 performing a psychoacoustic model analysis on frequency domain signal $S(f)$, to calculate a masking curve. Perceptually-weighting section 1307 estimates the perceptual importance level based on the masking curve, calculates respective weighting coefficients of the spectral coefficients, and then applies the respective weighting coefficients to the spectral coefficients.

[0106] Coding section 1308 encodes perceptually-weighted residual signal $S_{r_PW}(f)$ to generate a coding parameter.

[0107] Multiplexing section 1309 multiplexes the coding parameter with the signal to generate bit stream information, and transmits the bit stream information to the decoding side.

[0108] In speech decoding apparatus 1300B (FIG.13B), demultiplexing section 1310 demultiplexes the bit stream information to generate the coding parameter and LPC coefficients.

[0109] Decoding section 1311 decodes the coding parameter to generate decoded residual signal $S_{r_PW}(f)$ of a frequency domain.

[0110] LPC coefficient decoding section 1313 decodes the LPC coefficients.

[0111] Frequency-time transforming section 1312 transforms decoded residual signal $S_{r_PW}(f)$ of a frequency domain into a time domain using frequency-time transformation such as inverse discrete Fourier transform (IDFT) or inverse modified discrete cosine transform (IMDCT), to generate decoded residual signal $S_r(n)$ of a time domain.

[0112] Synthesis filter 1314 performs LPC synthesis filtering processing on decoded residual signal $S_r(n)$ of a time domain using the decoded LPC coefficients from LPC coefficient decoding section 1313, to obtain decoded time domain signal $\hat{S}(n)$.

[0113] FIG.14 illustrates a configuration of perceptually-weighting section 1307 according to the present example. FIG.14 illustrates a configuration to perceptually weight each spectral coefficient. Here, in FIG.14, the same components as in FIG.11 will be assigned the same reference numerals and detail explanations thereof will be omitted.

[0114] In perceptually-weighting section 1307, psychoacoustic model analyzing section 1401 calculates masking curve $M(f)$ based on spectral coefficient $S(f)$ of an original signal.

[0115] FIG.15 illustrates a state to perceptually weight each spectral coefficient.

[0116] As illustrated in FIG.15, energy levels of spectral coefficients $S(f_0)$, $S(f_1)$, $S(f_2)$, and $S(f_4)$ are lower than points of masking curve $M(f_0)$, $M(f_1)$, $M(f_2)$, and $M(f_4)$, respectively. Thus, the energy levels of these spectral coefficients are suppressed not to waste bits in encoding these spectral coefficients.

[0117] According to the present example, TCX coding determines the perceptual importance levels of the respective spectral coefficients, determines weighting coefficients according to the perceptual importance levels, applies the respective weighting coefficients to the spectral coefficients, and encodes the perceptually-weighted spectral coefficients.

[0118] By this means, the perceptually-weighting coefficients can calculate each spectral coefficient more accurately, in comparison with performing perceptual-weighting processing on a subband basis. Thus, it is possible to select and encode the most perceptually-important spectral coefficient and hence to obtain better coding performance.

[0119] According to the present example, only the coding side (speech coding apparatus 1300A) applies perceptually-weighting coefficients. That is, the decoding side (speech decoding apparatus 1300B) does not perform inverse weighting processing with respect to the above processing.

[0120] By this means, there is no need to transmit perceptually-weighting coefficients to a decoding side. Thus, it is possible to save bits in encoding the perceptually-weighting coefficients.

(embodiment of the invention)

[0121] FIG.16A illustrates a configuration of speech coding apparatus 1600A according to the present embodiment. FIG.16B also illustrates a configuration of speech decoding apparatus 1600B.

[0122] In the present embodiment, layer coding (scalable coding), in which a lower layer adopts a CELP coding and

a higher layer adopts a transform coding, perceptually weights each spectral coefficient. In the following explanation, although the layer coding including two layers of the lower layer and the higher layer will be explained as an example, it is possible to apply the present invention to the layer coding including three layers or more.

[0123] In speech coding apparatus 1600A (FIG. 16A), CELP coding section 1601 performs a CELP coding on an input signal so as to use redundancy of a signal in a time domain.

[0124] CELP decoding section 1602 generates synthesized signal $S_{syn}(n)$ using the CELP parameter.

[0125] By subtracting the synthesized signal from the input signal, subtractor 1612 obtains error signal $S_e(n)$ (error signal between the input signal and the synthesized signal).

[0126] Time-frequency transforming section 1604 transforms error signal $S_e(n)$ into frequency domain signal $S_e(f)$ (spectral coefficients) using time-frequency transformation such as discrete Fourier transform (DFT) or modified discrete cosine transform (MDCT).

[0127] Time-frequency transforming section 1603 transforms synthesized signal $S_{syn}(n)$ from CELP decoding section 1602 into frequency domain signal $S_{syn}(f)$ (spectral coefficients) using time-frequency transformation such as discrete Fourier transform (DFT) or modified discrete cosine transform (MDCT).

[0128] Perceptually-weighting section 1605 applies perceptual weighting of each spectral coefficient to spectral coefficient $S_e(f)$. Here, perceptually-weighting coefficients are calculated based on spectral coefficient $S_e(f)$ of an error signal and spectral coefficient $S_{syn}(f)$.

[0129] Coding section 1606 encodes the perceptually-weighted signal to generate a coding parameter.

[0130] Multiplexing section 1607 multiplexes the coding parameter and the CELP parameter to generate bit stream information and transmits the bit stream information to the decoding side.

[0131] In speech decoding apparatus 1600B (FIG. 16B), demultiplexing section 1608 demultiplexes the bit stream information to generate the coding parameter and CELP parameter.

[0132] Decoding section 1610 decodes the coding parameter to generate decoded error signal $S_e^{-}(f)$ of a frequency domain.

[0133] CELP decoding section 1609 generates synthesized signal $S_{syn}(n)$ using the CELP parameter.

[0134] Frequency-time transforming section 1611 transforms decoded residual signal $S_e^{-}(f)$ of a frequency domain into a time domain using frequency-time transformation such as inverse discrete Fourier transform (IDFT) or inverse modified discrete cosine transform (IMDCT), to generate decoded error signal $S_e^{-}(n)$ of a time domain.

[0135] By adding CELP synthesized signal $S_{syn}(n)$ and decoded error signal $S_e^{-}(n)$, adder 1613 generates decoded time domain signal $S^{-}(n)$.

[0136] FIG. 17 illustrates a configuration of perceptually-weighting section 1605 according to an example (configuration example 1). FIG. 17 illustrates a configuration to perceptually weight each spectral coefficient. Here, in FIG. 17, the same components as in FIG. 11 will be assigned the same reference numerals and detail explanations thereof will be omitted.

[0137] In perceptually-weighting section 1605 (configuration example 1) illustrated in FIG. 17, psychoacoustic model analyzing section 1701 calculates masking curve $M(f)$, based on spectral coefficient $S_{syn}(f)$ of the CELP decoded signal.

[0138] FIG. 18 illustrates a configuration of perceptually-weighting section 1605 according to the present embodiment (configuration example 2). FIG. 18 illustrates a configuration to perceptually weight each spectral coefficient.

[0139] In perceptually-weighting section 1605 (configuration example 2) illustrated in FIG. 18, adder 1805 generates spectrum $S(f)$ of the original signal, by adding spectrum $S_{syn}(f)$ of a CELP decoded signal and spectrum $S_e(f)$ of an error signal.

[0140] SNR calculating section 1801 calculates a signal-to-noise ratio of generated spectrum $S(f)$ of the original signal to spectrum $S_e(f)$ of the error signal. Signal-to-noise ratio $SNR(f)$ is calculated as the following equation.

$$SNR(f) = \frac{S^2(f)}{S_e^2(f)} \quad \dots \text{(Equation 5)}$$

[0141] Estimation section 1802 estimates perceptual importance level $pi(f)$ of each spectral coefficient, based on signal-to-noise ratio $SNR(f)$. Perceptual importance level $pi(f)$ is the parameter quantitatively representing how perceptually important the spectral coefficients are. Perceptual importance level $pi(f)$ showing a larger value means that the spectral coefficients corresponding to the $pi(f)$ are perceptually important. Perceptual importance $pi(f)$ is calculated based on signal-to-noise ratio $SNR(f)$ and energy of the spectral coefficients. The calculation may be performed in a logarithmic region, and, for example, perceptual importance level $pi(f)$ is calculated according to the following equation.

$$pi(f) = \log(S_e^2(f)) - \log(S_{ave}^2) + \log(SNR_{ave}) - \log(SNR(f)) \dots \text{(Equation 6)}$$

[0142] Here, S_{ave}^2 represents the average energy of spectral coefficients included a subband, and is calculated as the following equation.

$$S_{ave}^2 = \frac{\sum_{i=0}^{N-1} S_e^2(i)}{N} \dots \text{(Equation 7)}$$

[0143] Also, SNR_{ave} represents the signal-to-noise ratio of the entire spectral coefficients included the subband, and is calculated as the following equation.

$$SNR_{ave} = \frac{\sum_{i=0}^{N-1} S^2(i)}{\sum_{i=0}^{N-1} S_e^2(i)} \dots \text{(Equation 8)}$$

[0144] Perceptual importance level $pi(f)$ may be calculated as the following equation using terms of a signal-to-noise ratio.

$$pi(f) = \log(SNR_{ave}) - \log(SNR(f)) \dots \text{(Equation 9)}$$

[0145] Weighting coefficient calculating section 1803 calculates weighting coefficient $W(f)$, based on perceptual importance level $pi(f)$. Weighting coefficient $W(f)$ is used for applying a weight to spectral coefficient $S(f)$. As perceptual importance level $pi(f)$ shows a larger value, weighting coefficient $W(f)$ becomes a larger value. For example, weighting coefficient $W(f)$ is calculated as the following equation.

$$W(f) = e^{pi(f)} \dots \text{(Equation 10)}$$

[0146] Weighting section 1804 multiplies spectral coefficient $S(f)$ by weighting coefficient $W(f)$ to generate perceptually-weighted spectral coefficient $S_{e_PW}(f)$. Thus, spectral coefficient $S_{e_PW}(f)$ is calculated as the following equation.

$$S_{e_PW}(f) = W(f) * S_e(f) \dots \text{(Equation 11)}$$

[0147] FIG. 19 illustrates a state where each spectral coefficient is perceptually weighted.

[0148] Focusing on spectral coefficient $S(f1)$ in FIG. 19 shows that this spectral coefficient has a larger amplitude value than other spectral coefficients. Also, signal-to-noise ratio $SNR(f1)$ at frequency $f1$ is a maximum value in comparison with other signal-to-noise ratios. At this time, the present embodiment multiplies a small weighting coefficient $W(f1)$ which is less than 1 to spectral coefficient $S_e(f1)$ of an error signal, and hence the amplitude value of weighted spectral coefficient $S_{e_PW}(f1)$ becomes smaller than that of $S_e(f1)$.

[0149] As an example, when perceptual importance level $pi(f)$ and weighting coefficient $W(f)$ are calculated as the above, perceptually-weighted spectral coefficient $S_{e_PW}(f1)$ is represented as the following equation, to reveal that $S_{e_PW}(f1)$ becomes lower than spectral coefficient $S_e(f1)$.

$$\begin{aligned}
 &pi(f1) = \log(SNR_{ave}) - \log(SNR(f1)) < 0; \\
 &then \\
 &W(f1) = e^{pi(f1)} < 1; \\
 &then \\
 &S_{e_PW}(f1) = W(f1) * S_e(f1) < S_e(f1); \quad \dots \text{(Equation 12)}
 \end{aligned}$$

[0150] As described above, by calculating weighting coefficients on a frequency basis according to the signal-to-noise ratio, the present embodiment lowers the importance of the spectrum with a high signal-to-noise ratio to set coding bits less likely to be distributed to this spectrum.

[0151] As a result, distribution of more coding bits to other spectra with low signal-to-noise ratios improves the sound quality.

[0152] Embodiments of the present invention have been described above.

[0153] Although a case has been described with the above embodiments as an example where the present invention is implemented with hardware, the present invention can be implemented with software.

[0154] Each function block employed in the description of each of the aforementioned embodiments may typically be implemented as an LSI constituted by an integrated circuit. These may be individual chips or partially or totally contained on a single chip. "LSI" is adopted here but this may also be referred to as "IC," "system LSI," "super LSI," or "ultra LSI" depending on differing extents of integration.

[0155] The method of implementing integrated circuitry is not limited to LSI, and implementation by means of dedicated circuitry or a general-purpose processor may also be used. After LSI manufacture, utilization of a programmable FPGA (Field Programmable Gate Array) or a reconfigurable processor where connections and settings of circuit cells within an LSI can be reconfigured is also possible.

[0156] In the event of the introduction of an integrated circuit implementation technology whereby LSI is replaced by a different technology as an advance in, or derivation from, semiconductor technology, integration of the function blocks may of course be performed using that technology. Application of biotechnology is also possible.

Industrial Applicability

[0157] The present invention is suitable for a communication apparatus encoding speech.

Reference Signs List

[0158]

- 1000A Speech coding apparatus
- 1000B Speech decoding apparatus
- 1001 Time-frequency transforming section
- 1002 Psychoacoustic model analyzing section
- 1003 Perceptually-weighting section
- 1004 Coding section
- 1005 Multiplexing section
- 1006 Demultiplexing section
- 1007 Decoding section
- 1008 Frequency-time transforming section
- 1101 Estimation section
- 1102 Weighting coefficient calculating section
- 1103 Weighting section
- 1300A Speech coding apparatus
- 1300B Speech decoding apparatus
- 1301 LPC analyzing section
- 1302 Coding section
- 1303 Decoding section
- 1304 Inverse filter section
- 1305 Time-frequency transforming section
- 1306 Time-frequency transforming section

	1307 Perceptually-weighting section
	1308 Coding section
	1309 Multiplexing section
	1310 Demultiplexing section
5	1311 Decoding section
	1312 Frequency-time transforming section
	1313 LPC coefficient decoding section
	1314 Synthesis filter
	1401 Psychoacoustic model analyzing section
10	1600A Speech coding apparatus
	1600B Speech decoding apparatus
	1601 CELP coding section
	1602 CELP decoding section
	1603 Time-frequency transforming section
15	1604 Time-frequency transforming section
	1605 Perceptually-weighting section
	1606 Coding section
	1607 Multiplexing section
	1608 Demultiplexing section
20	1609 CELP decoding section
	1610 Decoding section
	1611 Frequency-time transforming section
	1612 Subtractor
	1613 Adder
25	1701 Psychoacoustic model analyzing section
	1801 SNR calculating section
	1802 Estimation section
	1803 Weighting coefficient calculating section
	1804 Weighting section
30	1805 Adder

Claims

- 35 1. A speech coding apparatus (1600A) adapted to perform layer coding including at least two layers of a lower layer and a higher layer, the speech coding apparatus comprising:
- a generating section (1612) adapted to generate an error signal between a decoded signal of the lower layer and an input signal;
- 40 an estimation section (1801) adapted to calculate a signal-to-noise ratio using the input signal and the error signal and adapted to estimate (1802) respective perceptual importance levels of a plurality of spectral coefficients of different frequencies in the error signal, based on the signal-to-noise ratio;
- a calculating section (1803) adapted to calculate respective weighting coefficients of the plurality of spectral coefficients based on the respective estimated importance levels;
- 45 a weighting section (1804) adapted to weight each of the plurality of spectral coefficients using the respective calculated weighting coefficients; and
- a coding section (1606) adapted to encode the plurality of weighted spectral coefficients.
- 50 2. A speech coding method for performing layer coding including at least two layers of a lower layer and a higher layer, the speech coding method comprising the steps of:
- generating an error signal between a decoded signal of the lower layer and an input signal;
- calculating a signal-to-noise ratio using the input signal and the error signal and estimating respective perceptual importance levels of a plurality of spectral coefficients of different frequencies in the error signal, based on the
- 55 signal-to-noise ratio;
- calculating respective weighting coefficients of the plurality of spectral coefficients based on the respective estimated importance levels;
- weighting each of the plurality of spectral coefficients using the respective calculated weighting coefficients; and

encoding the plurality of weighted spectral coefficients.

Patentansprüche

1. Sprachcodierungsvorrichtung (1600A), die ausgebildet ist, eine Schichtcodierung mit mindestens zwei Schichten mit einer tieferen Schicht und einer höheren Schicht auszuführen, wobei die Sprachcodierungsvorrichtung umfasst:

einen Erzeugungsabschnitt (1612), der ausgebildet ist, ein Fehlersignal zwischen einem decodierten Signal der tieferen Schicht und einem Eingangssignal zu erzeugen;
 einen Abschätzabschnitt (1801), der ausgebildet ist, ein Signal-zu-Rauschen-Verhältnis unter Anwendung des Eingangssignals und des Fehlersignals zu berechnen, und der ferner ausgebildet ist, entsprechende Wahrnehmungswichtungsstufen mehrerer spektraler Koeffizienten unterschiedlicher Frequenzen in dem Fehlersignal auf der Grundlage des Signal-zu-Rauschen-Verhältnisses abzuschätzen;
 einen Berechnungsabschnitt (1803), der ausgebildet ist, jeweilige Gewichtungskoeffizienten der mehreren spektralen Koeffizienten auf der Grundlage der jeweiligen abgeschätzten Wichtungsstufen zu berechnen;
 einen Gewichtungsabschnitt (1804), der ausgebildet ist, jeden der mehreren spektralen Koeffizienten unter Anwendung der jeweiligen berechneten Gewichtungskoeffizienten zu gewichten; und
 einen Codierabschnitt (1906), der ausgebildet ist, die mehreren gewichteten spektralen Koeffizienten zu codieren.

2. Sprachcodierungsverfahren zur Ausführung einer Schichtcodierung mit mindestens zwei Schichten mit einer tieferen Schicht und einer höheren Schicht, wobei das Sprachcodierungsverfahren die Schritte umfasst:

Erzeugen eines Fehlersignals zwischen dem decodierten Signal der tieferen Schicht und einem Eingangssignal;
 Berechnen eines Signal-zu-Rauschen-Verhältnisses unter Anwendung des Eingangssignals und des Fehlersignals und Abschätzen entsprechender Wahrnehmungswichtungsstufen mehrerer spektraler Koeffizienten unterschiedlicher Frequenzen in dem Fehlersignal auf der Grundlage des Signal-zu-Rauschen-Verhältnisses;
 Berechnen jeweiliger Gewichtungskoeffizienten der mehreren spektralen Koeffizienten auf der Grundlage der jeweiligen abgeschätzten Wichtungsstufen;
 Gewichten jedes der mehreren spektralen Koeffizienten unter Anwendung der jeweiligen berechneten Gewichtungskoeffizienten; und
 Codieren der mehreren gewichteten spektralen Koeffizienten.

Revendications

1. Appareil de codage vocal (1600A) adapté pour mettre en oeuvre un codage par couches incluant au moins deux couches parmi une couche inférieure et une couche supérieure, l'appareil de codage vocal comprenant :

une section de génération (1612) adaptée pour générer un signal d'erreur entre un signal décodé de la couche inférieure et un signal d'entrée ;
 une section d'estimation (1801) adaptée pour calculer un rapport signal sur bruit en utilisant le signal d'entrée et le signal d'erreur et adaptée pour estimer (1802) des niveaux d'importance perceptuelle respectifs d'une pluralité de coefficients spectraux de différentes fréquences dans le signal d'erreur, sur base du rapport signal sur bruit ;
 une section de calcul (1803) adaptée pour calculer des coefficients de pondération respectifs de la pluralité de coefficients spectraux sur base des niveaux d'importance estimés respectifs ;
 une section de pondération (1804) adaptée pour pondérer chaque coefficient de la pluralité de coefficients spectraux en utilisant les coefficients de pondération calculés respectifs ; et
 une section de codage (1606) adaptée pour coder la pluralité de coefficients spectraux pondérés.

2. Procédé de codage vocal pour mettre en oeuvre un codage par couches incluant au moins deux couches parmi une couche inférieure et une couche supérieure, le procédé de codage vocal comprenant les étapes suivantes :

génération d'un signal d'erreur entre un signal décodé de la couche inférieure et un signal d'entrée ;
 calcul d'un rapport signal sur bruit en utilisant le signal d'entrée et le signal d'erreur et estimation des niveaux d'importance perceptuelle respectifs d'une pluralité de coefficients spectraux de différentes fréquences dans

EP 2 525 355 B1

le signal d'erreur, sur base du rapport signal sur bruit ;
calcul des coefficients de pondération respectifs de la pluralité de coefficients spectraux sur base des niveaux d'importance estimés respectifs ;
pondération de chaque coefficient de la pluralité de coefficients spectraux en utilisant les coefficients de pondération calculés respectifs ; et
codage de la pluralité de coefficients spectraux pondérés.

5

10

15

20

25

30

35

40

45

50

55

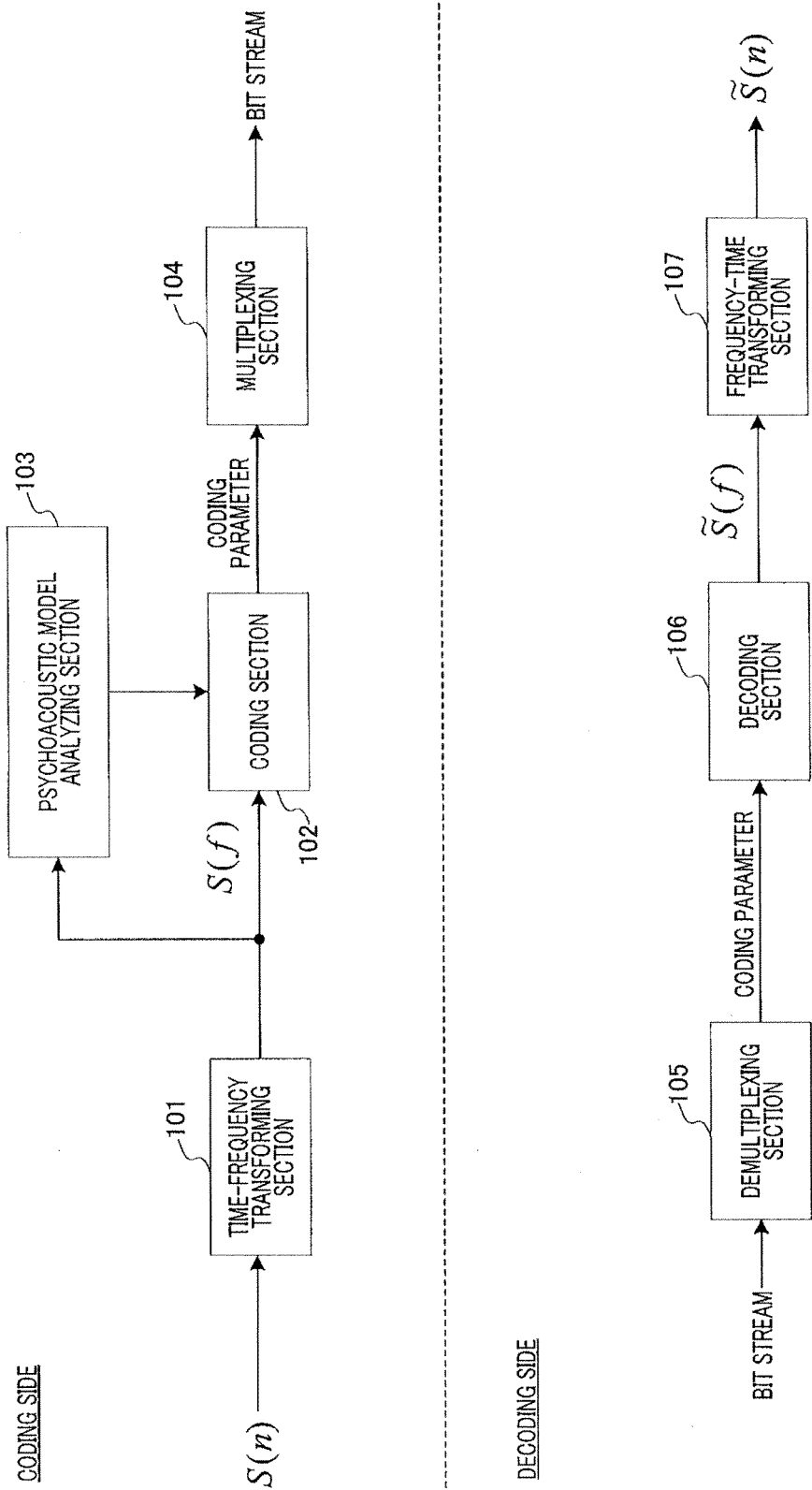


FIG.1

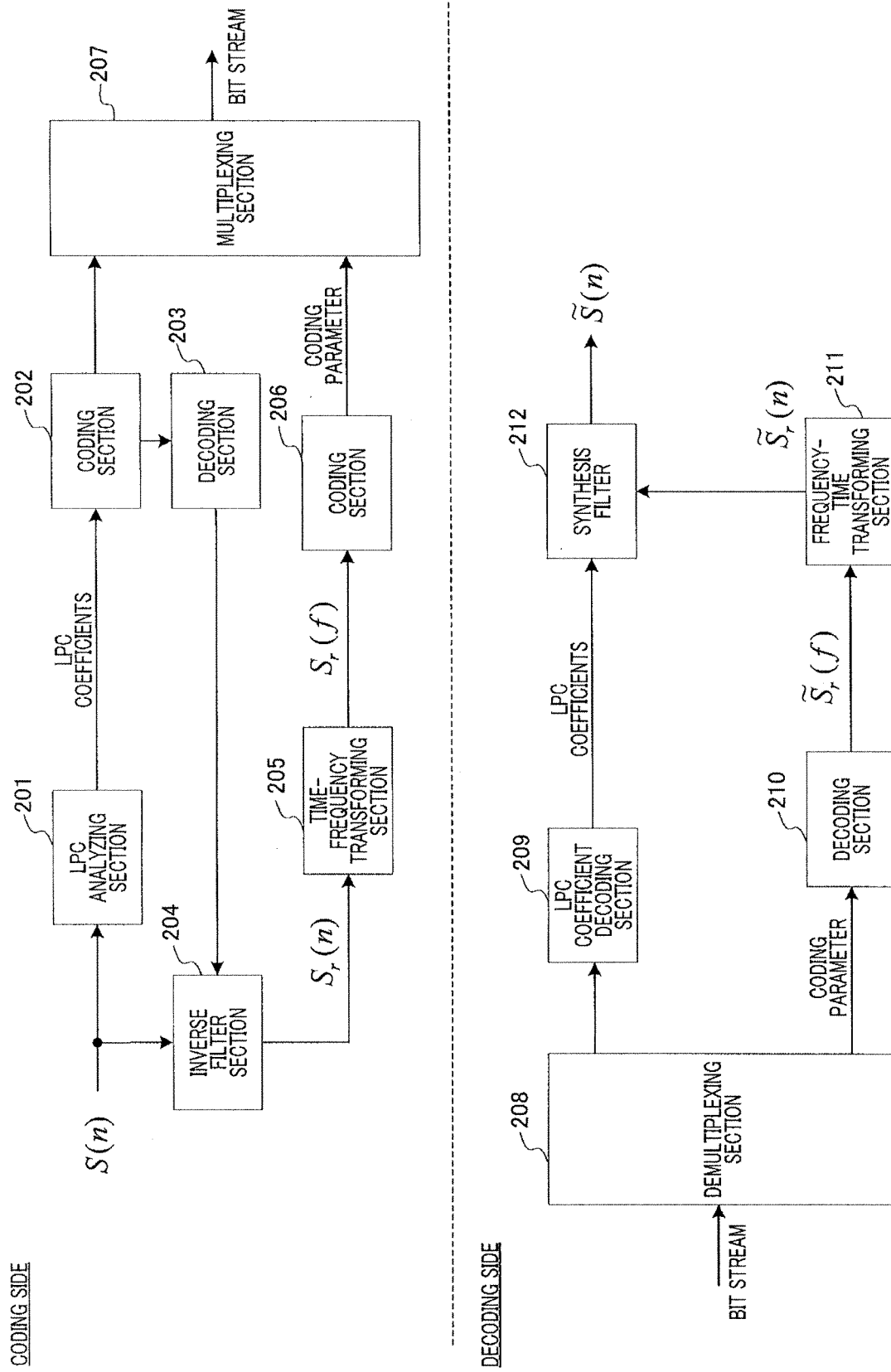


FIG.2

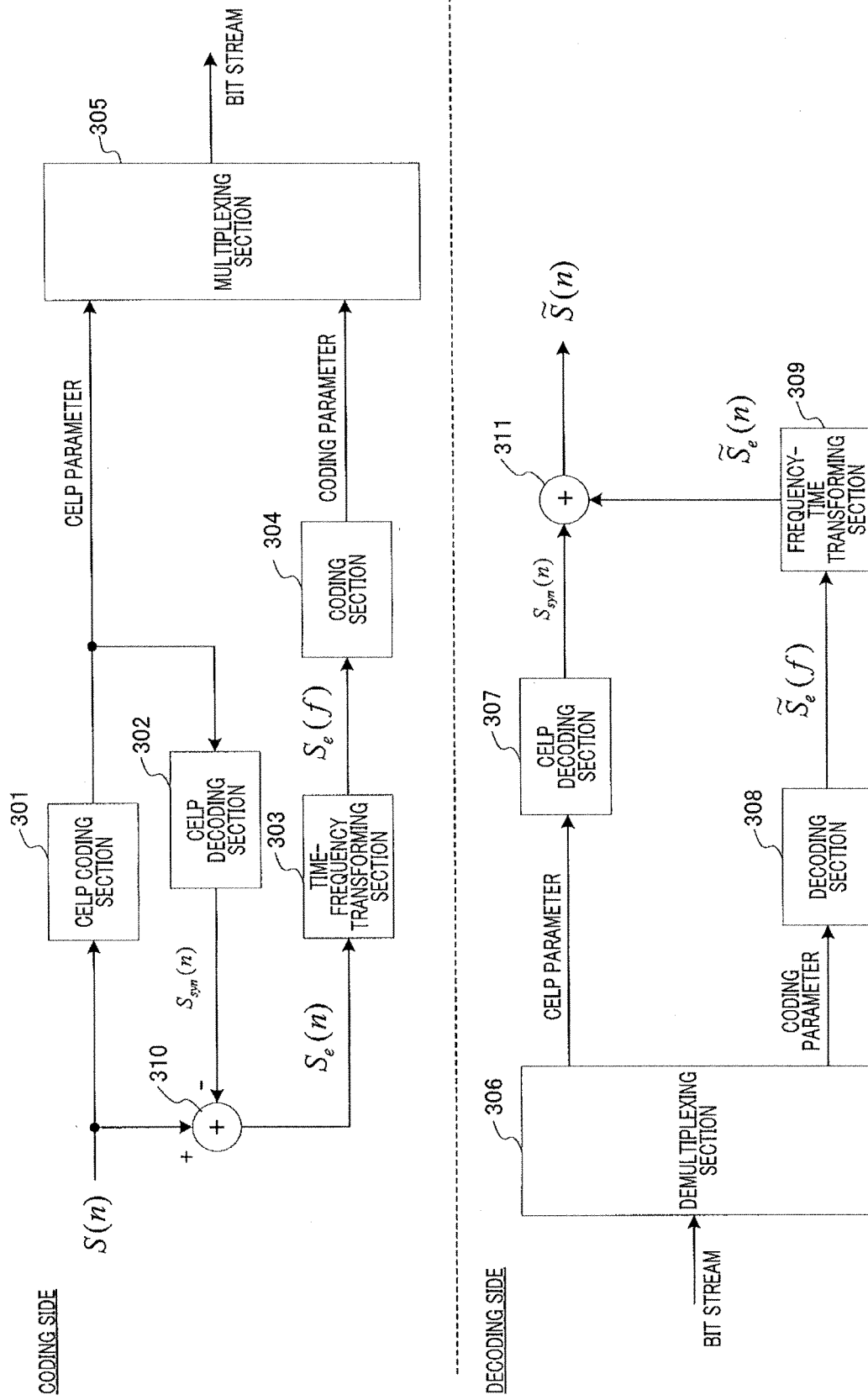
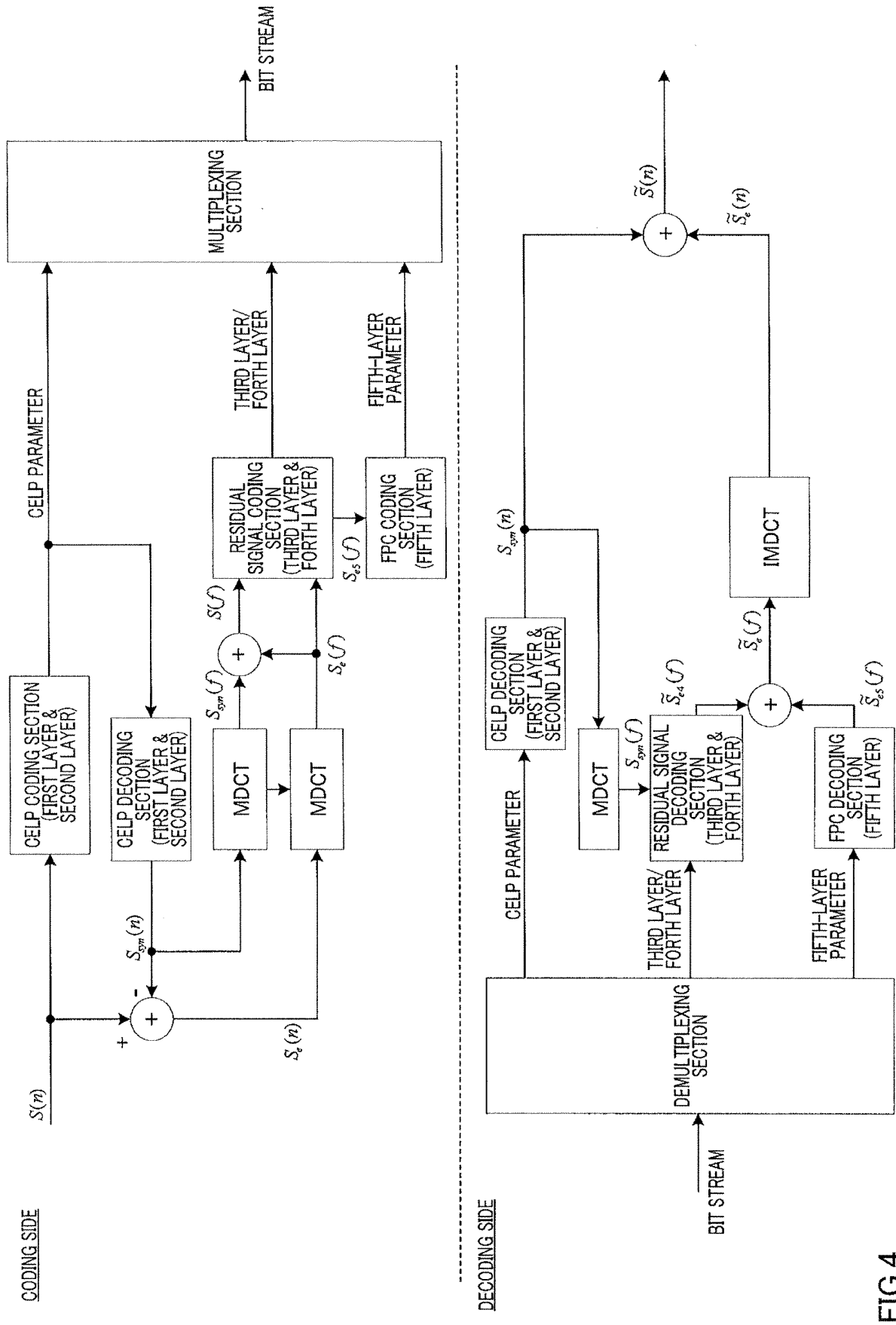


FIG.3



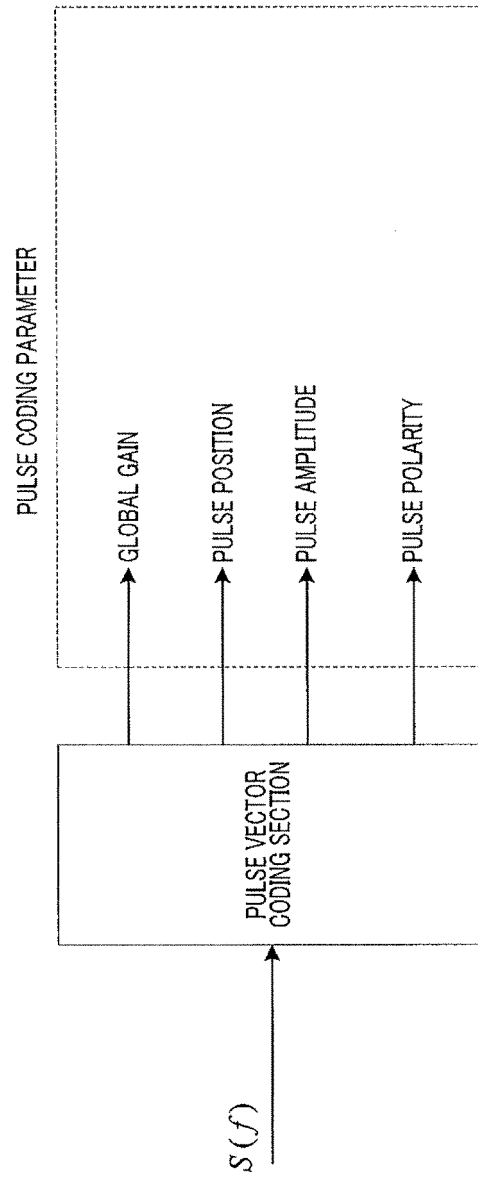


FIG.5

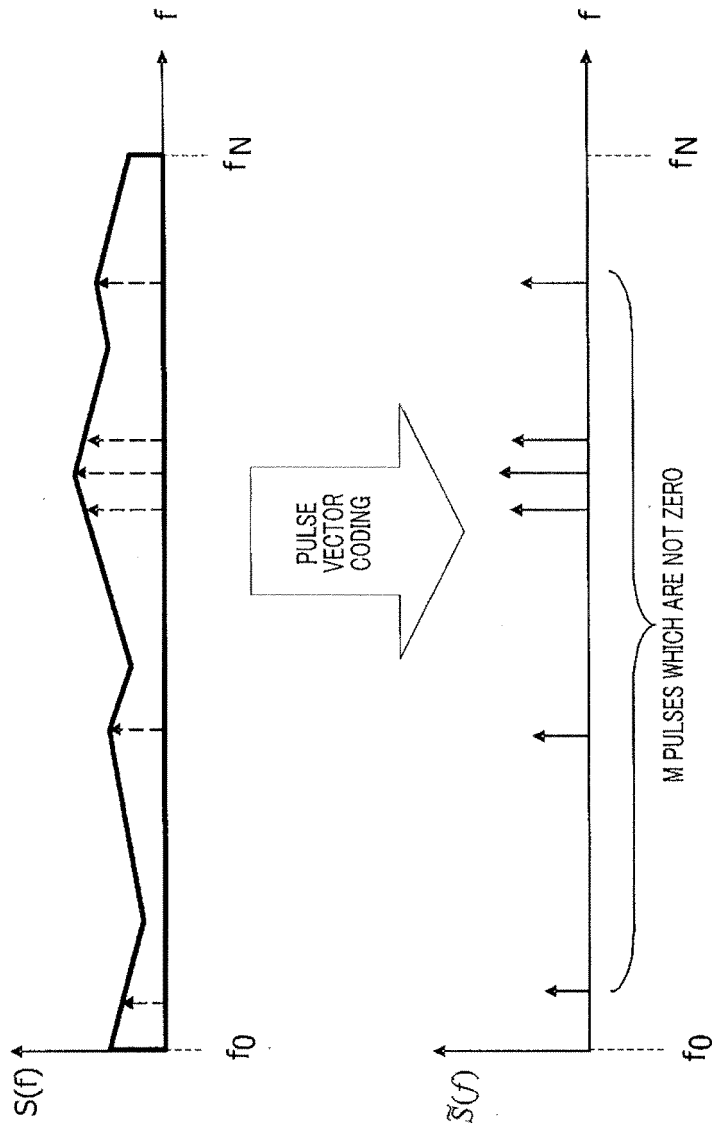


FIG.6

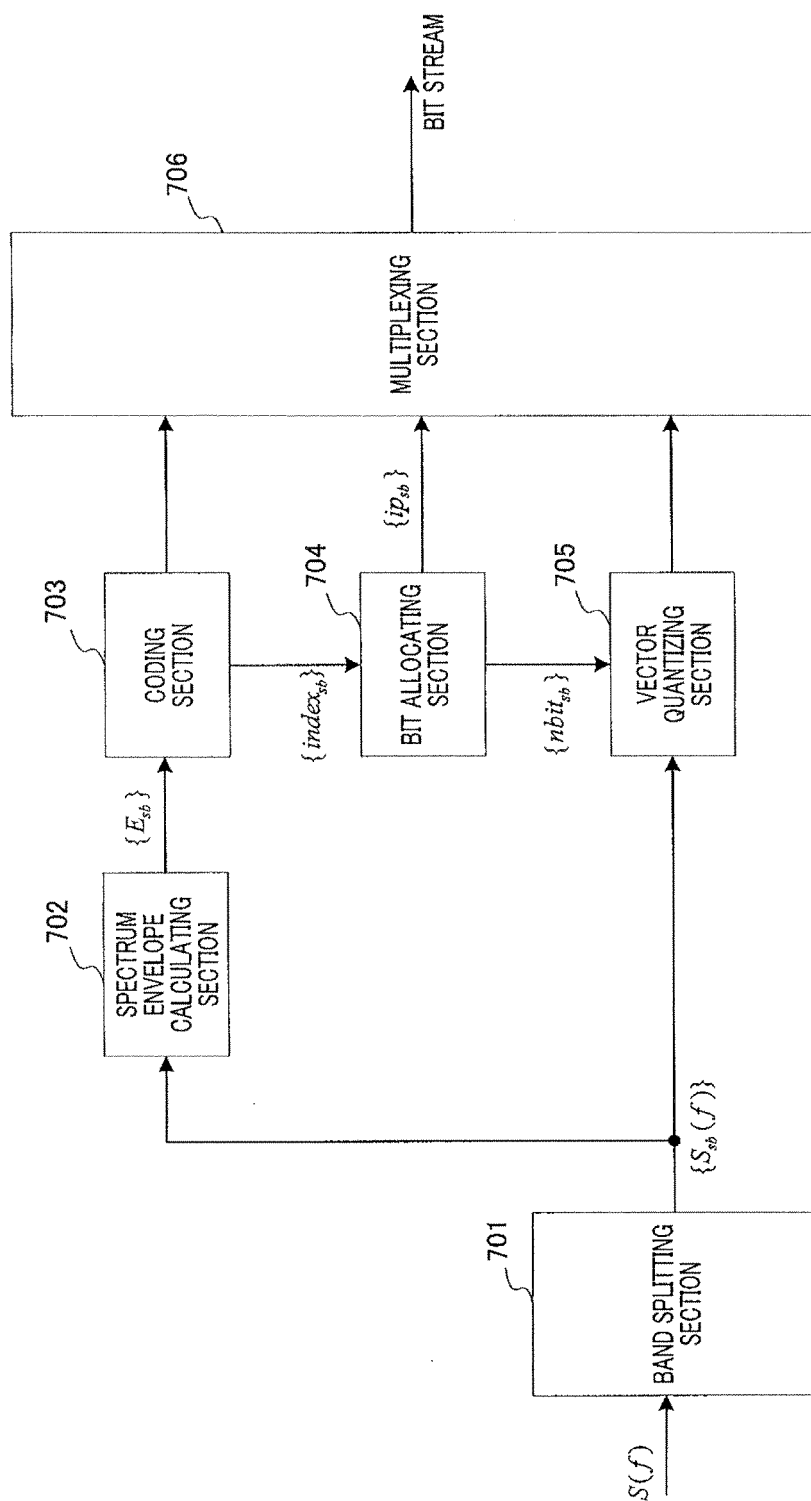


FIG. 7

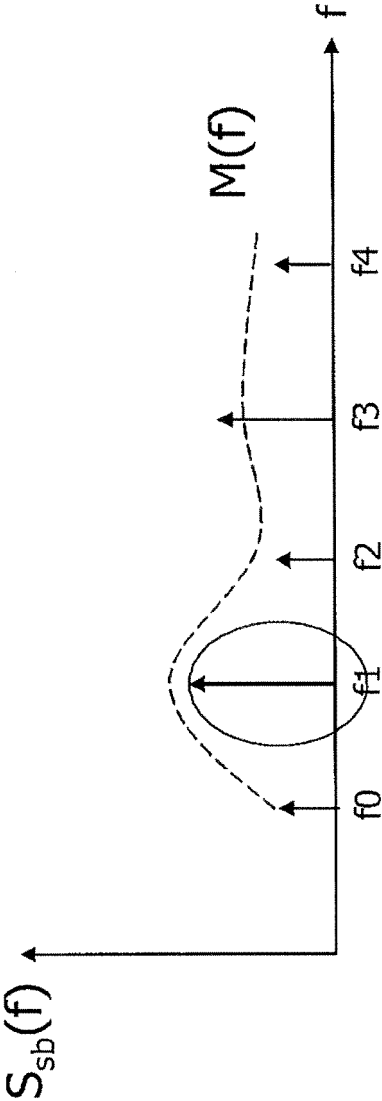


FIG.8

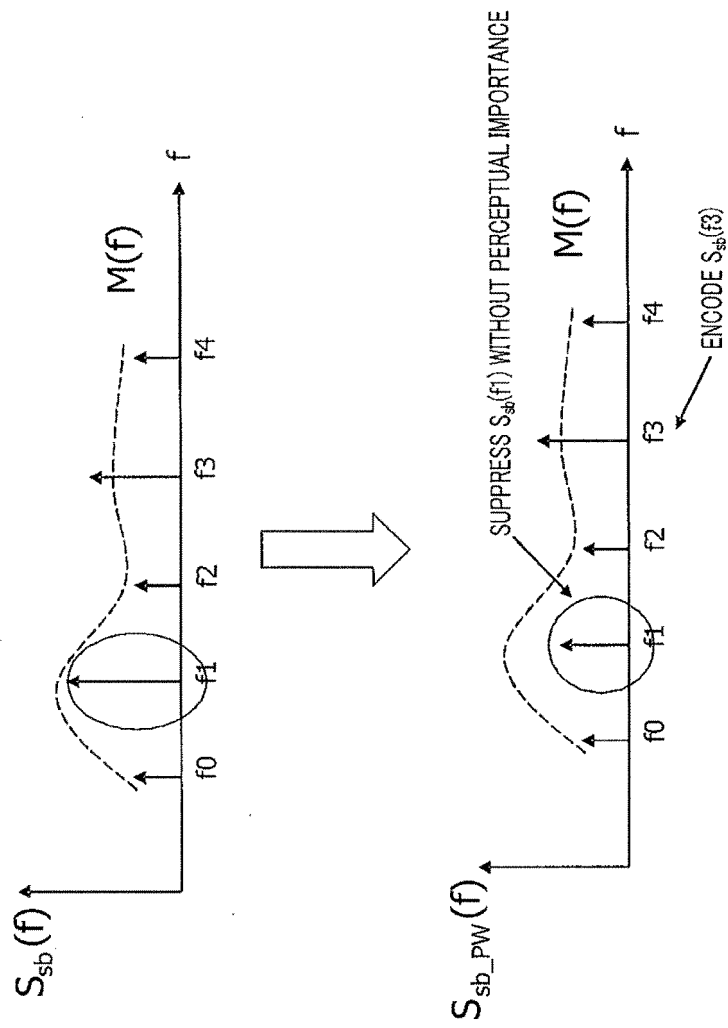


FIG.9

1000A

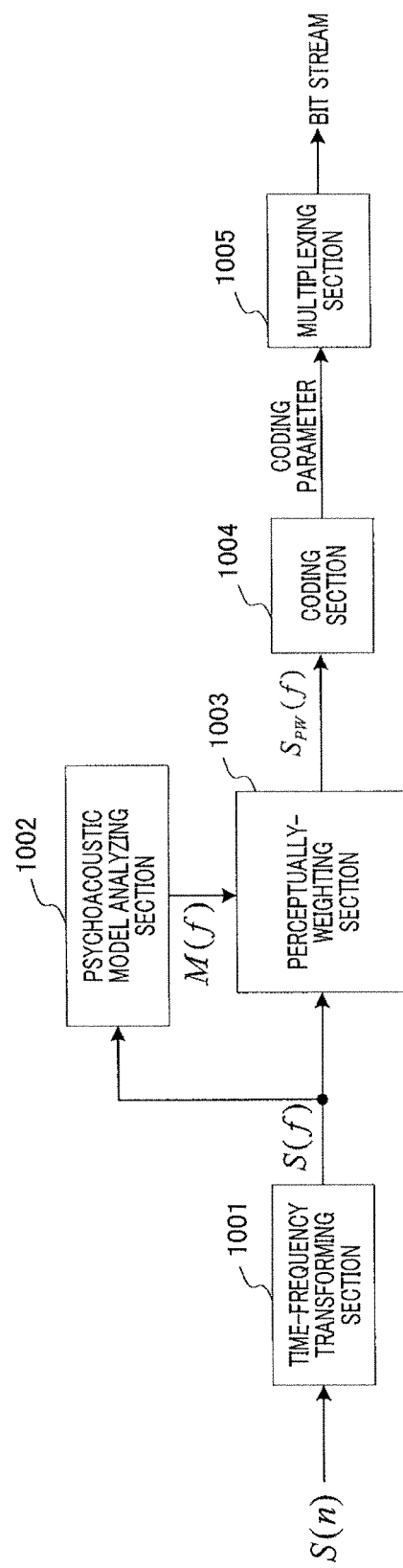


FIG.10A

1000B

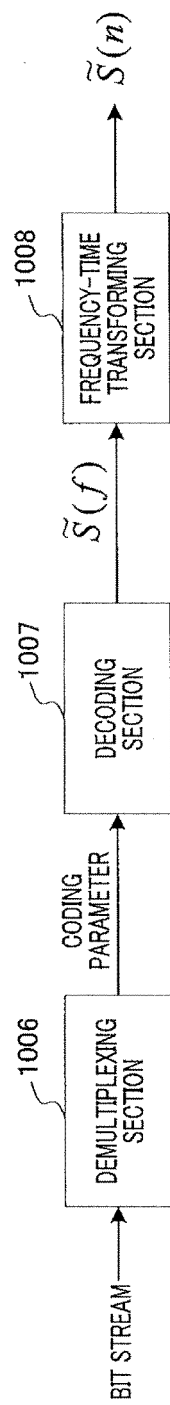


FIG.10B

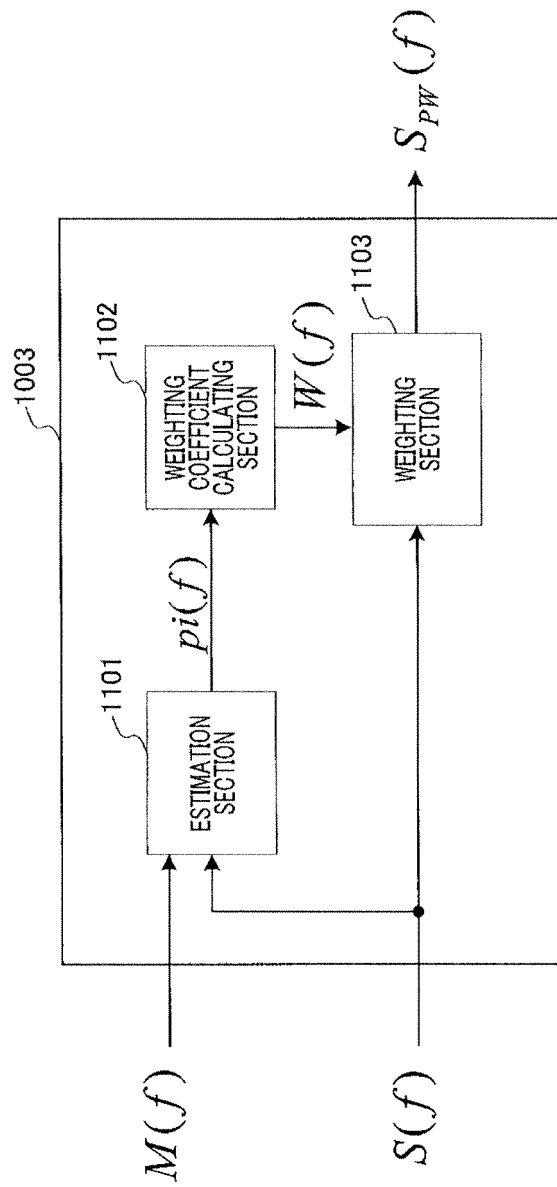


FIG.11

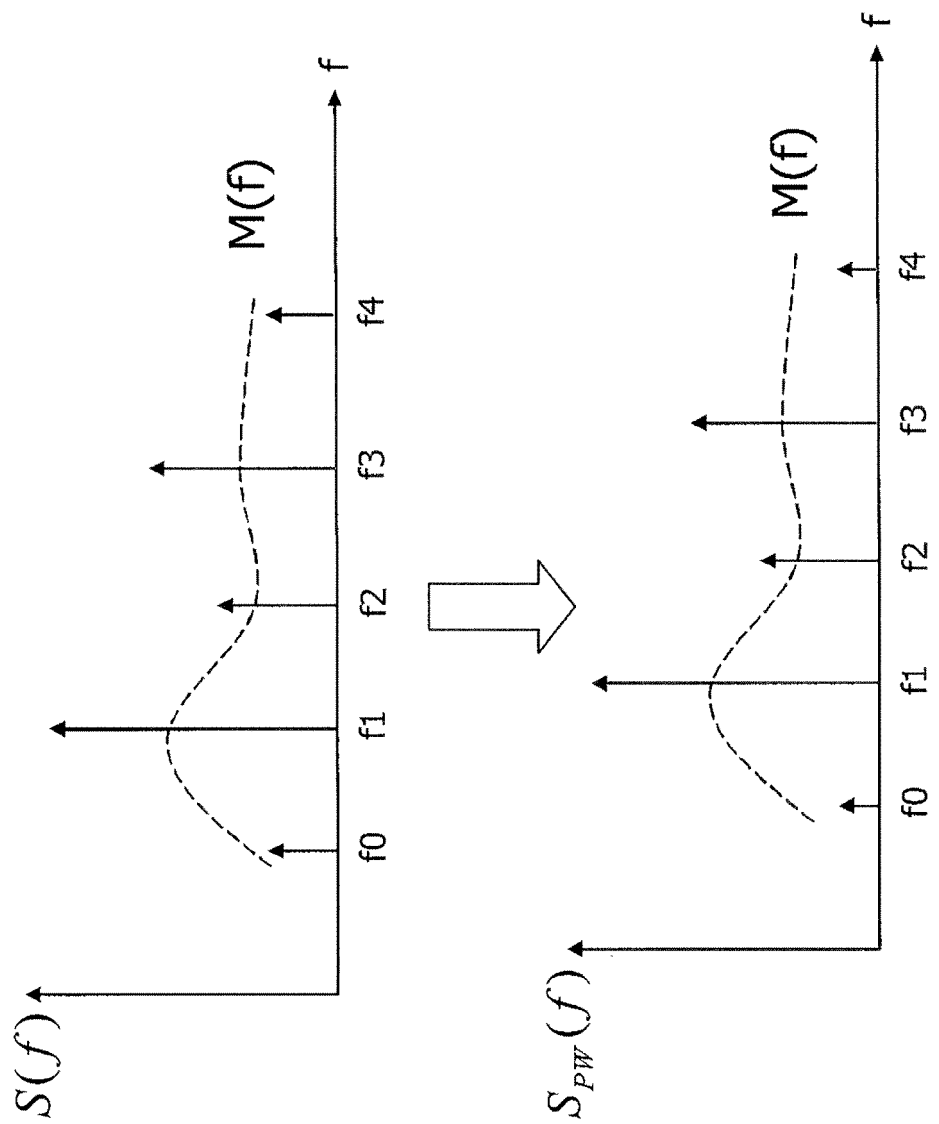


FIG.12

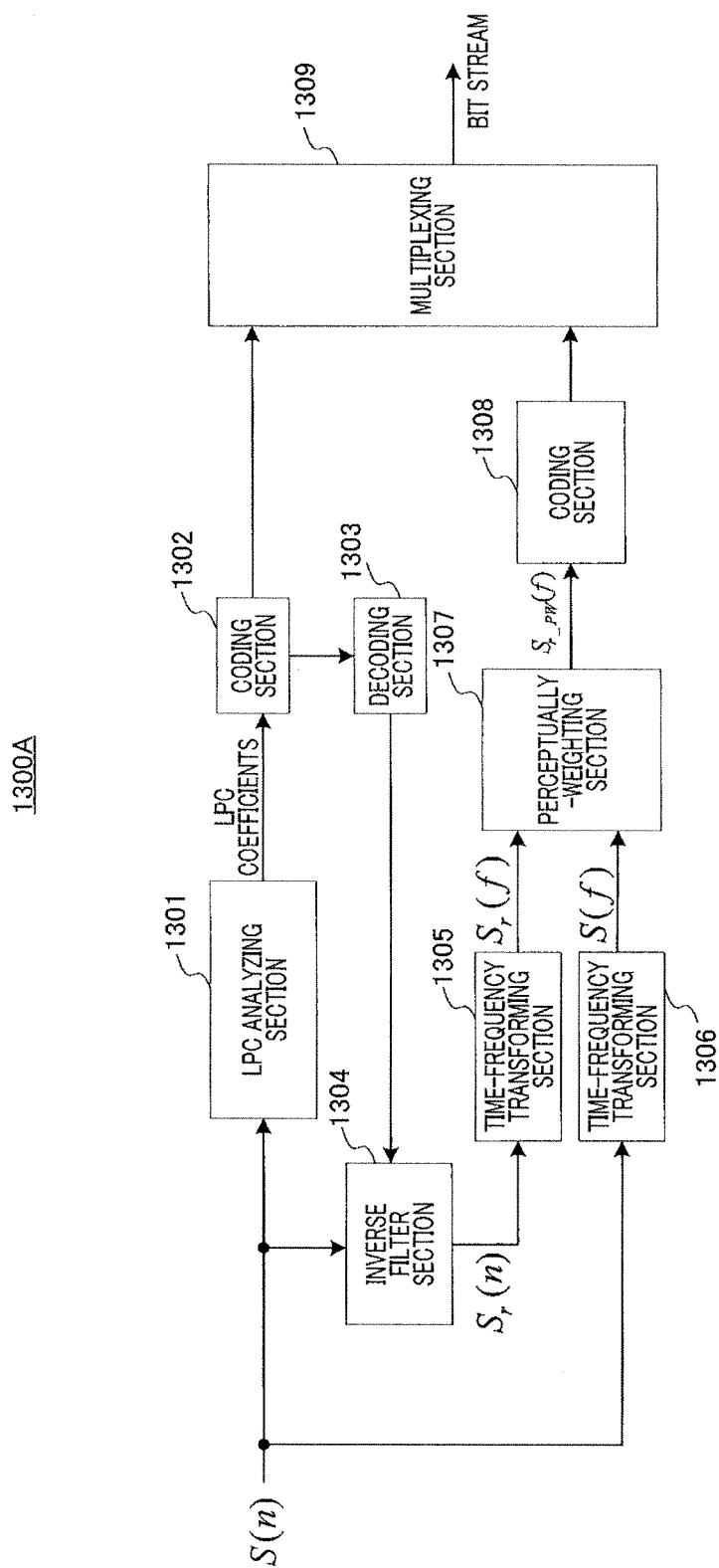


FIG.13A

1300B

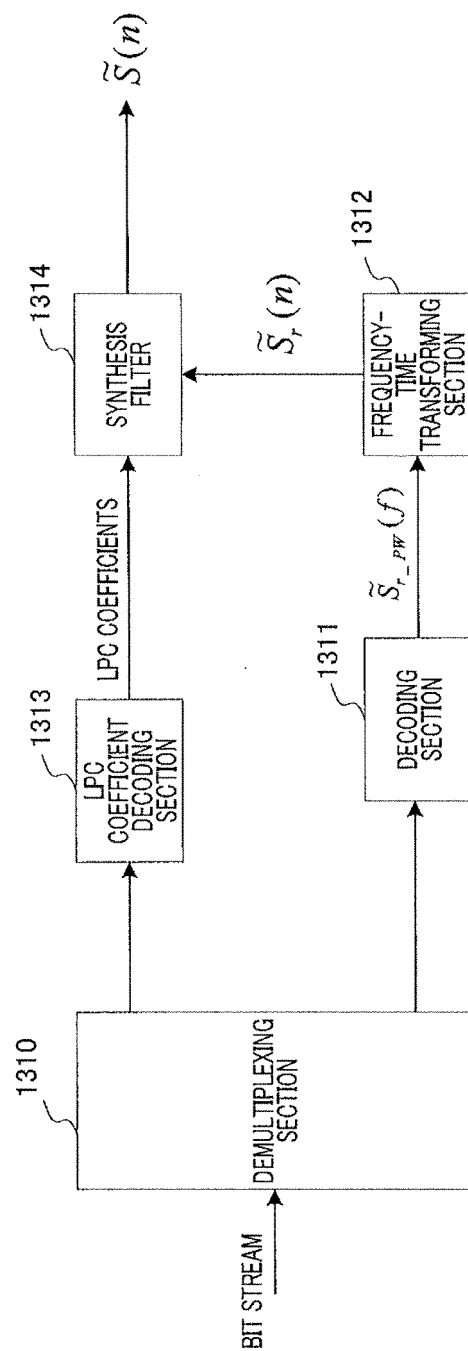


FIG.13B

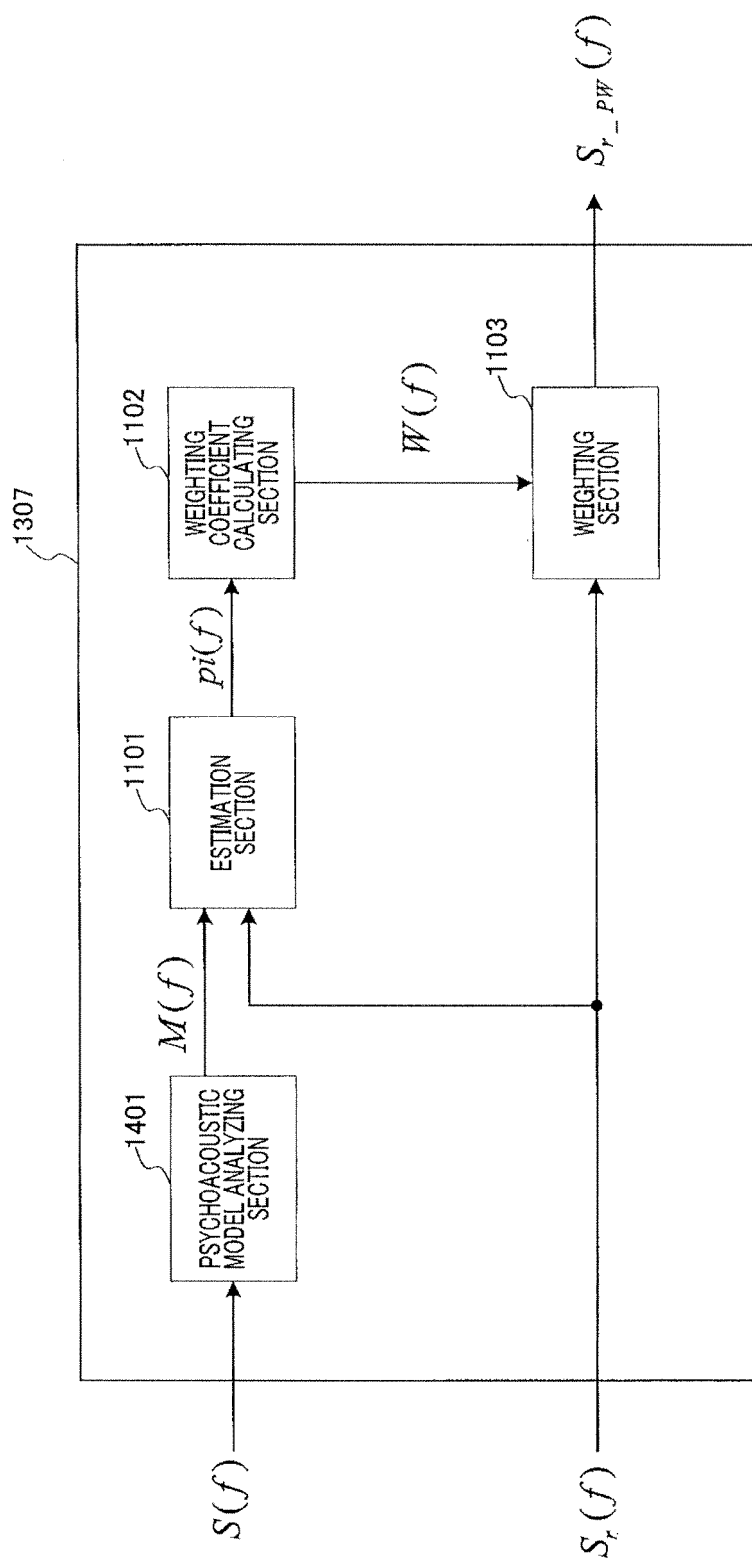


FIG.14

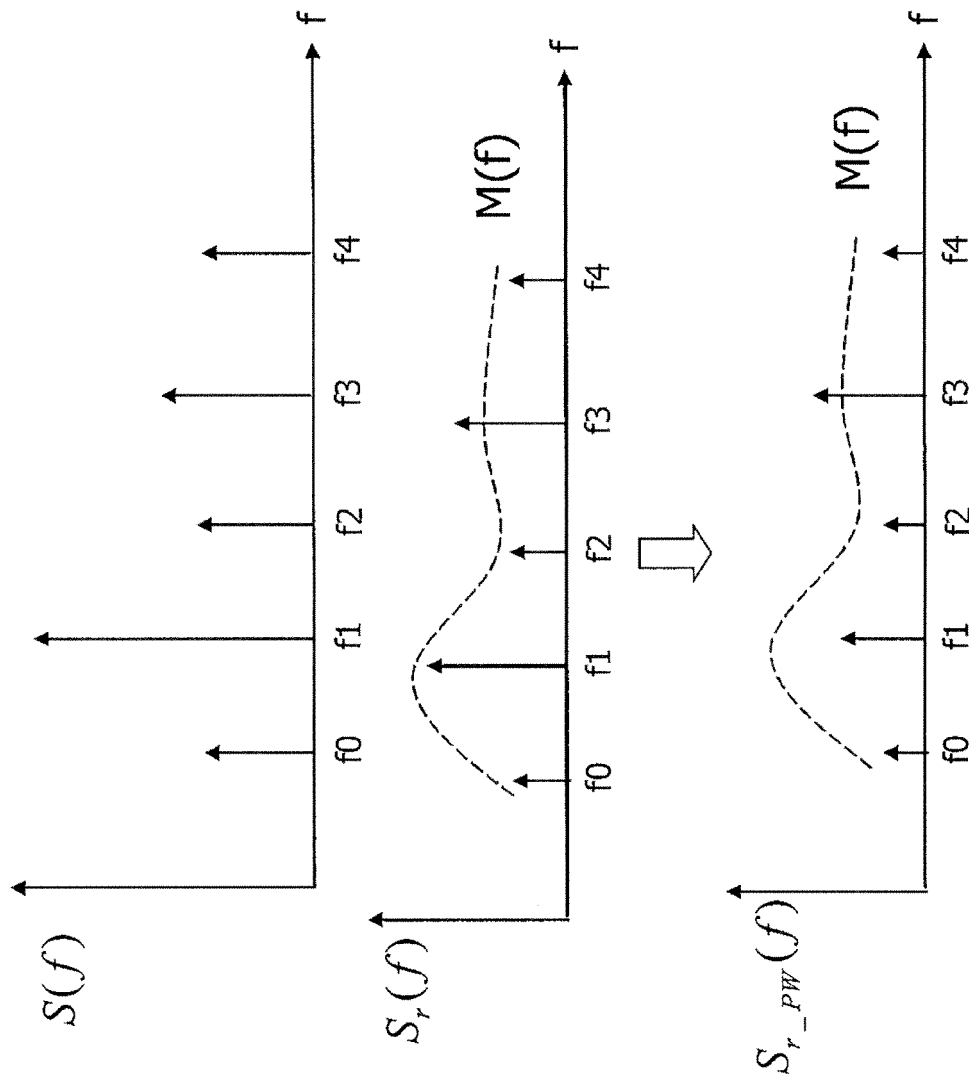


FIG.15

1600A

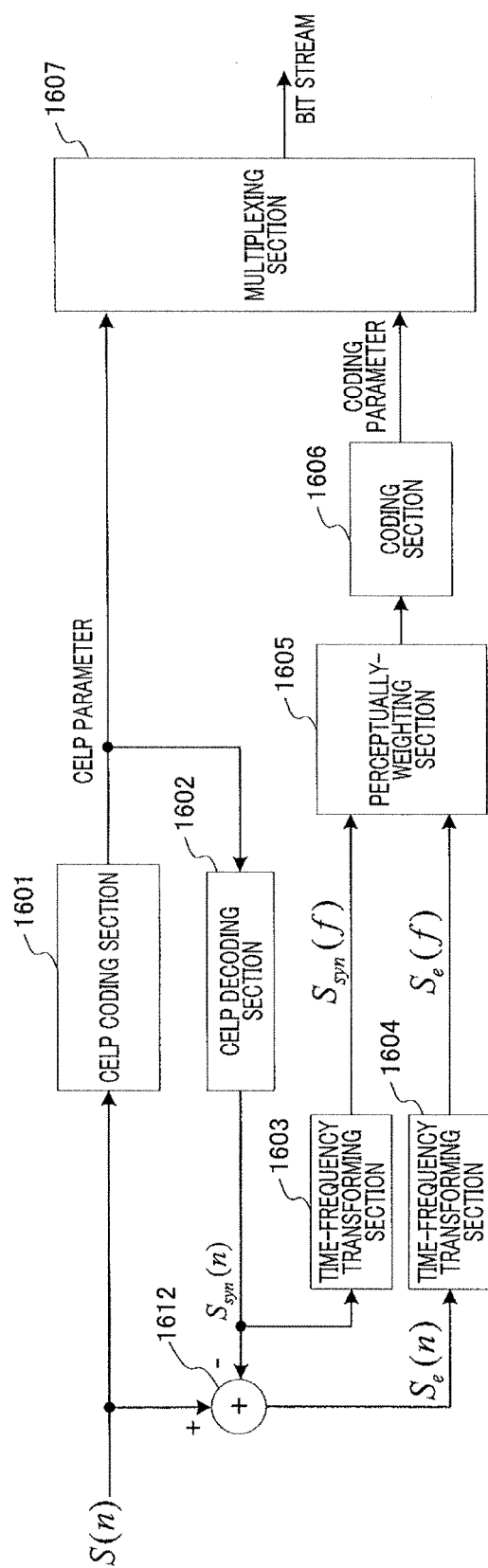


FIG.16A

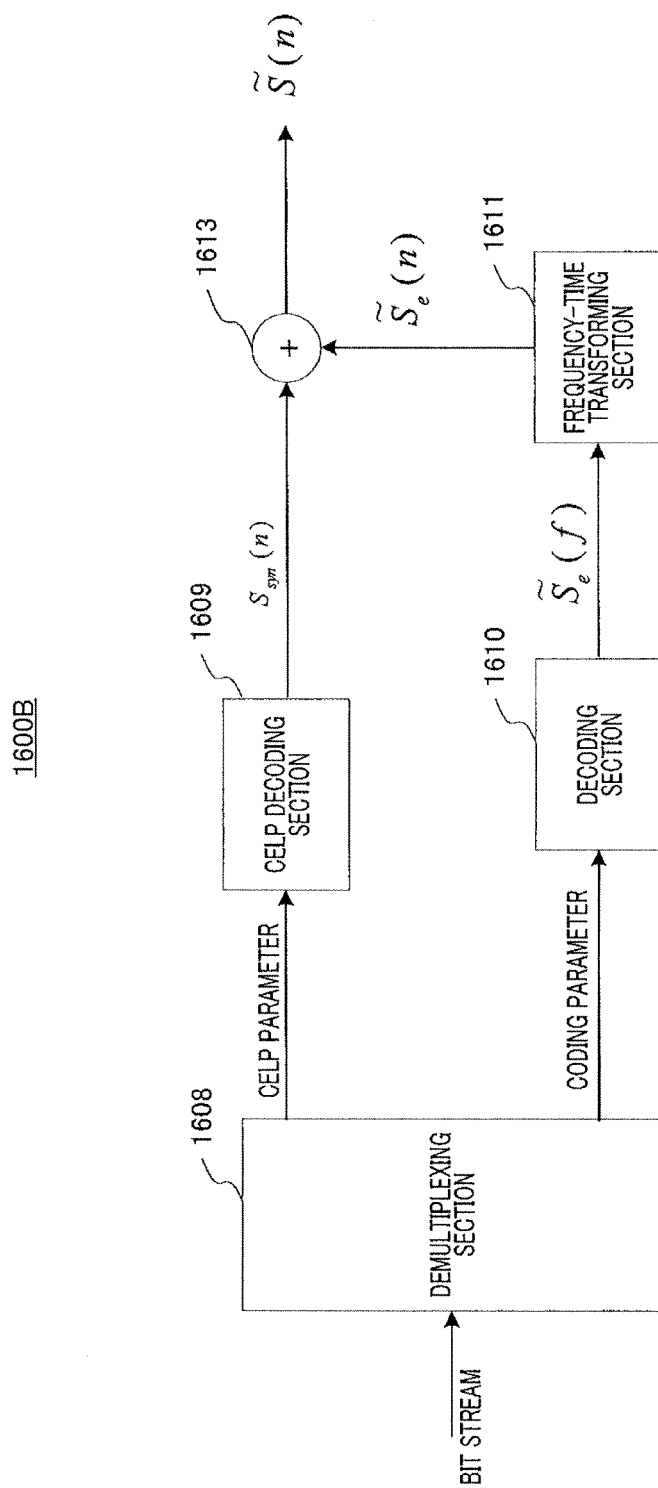


FIG.16B

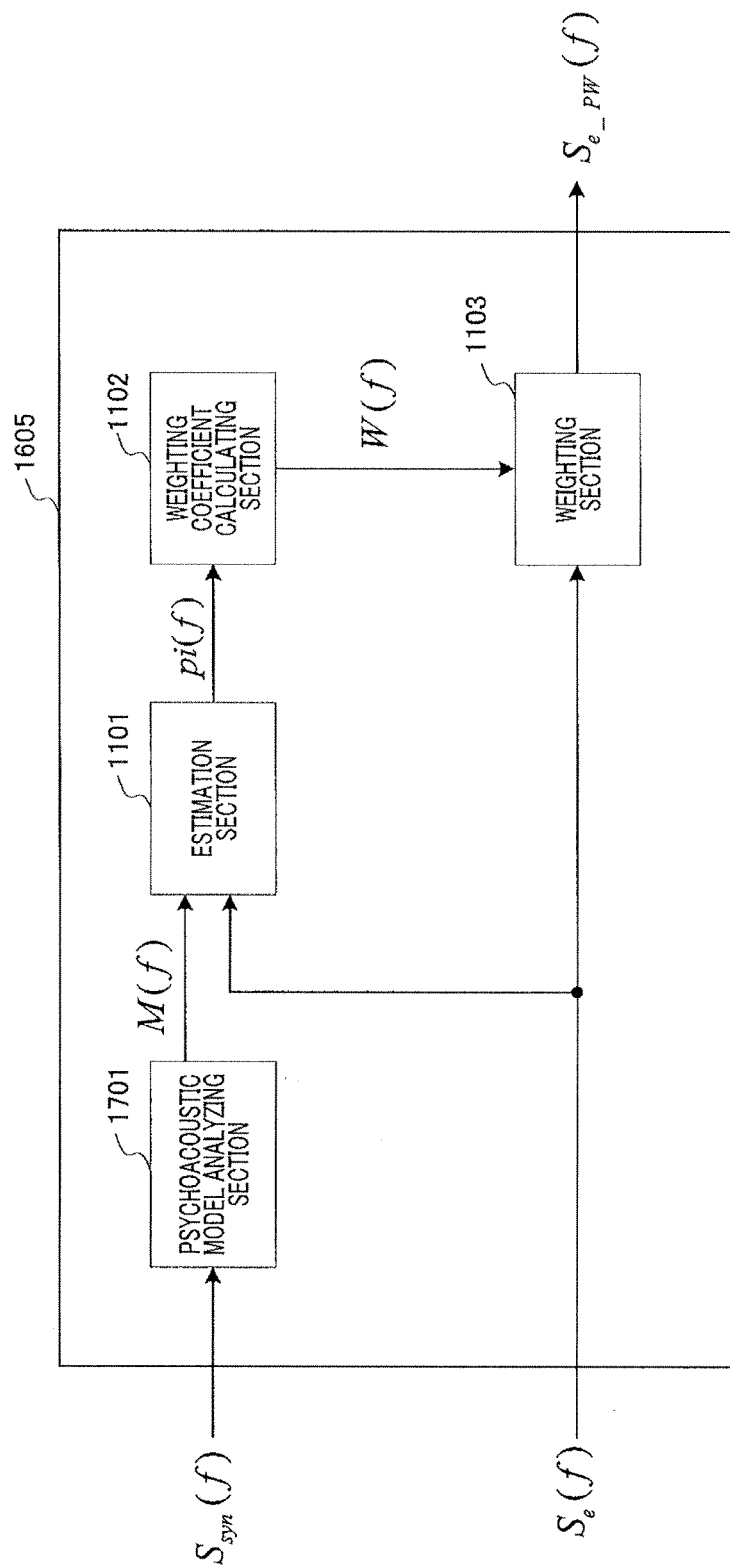


FIG.17

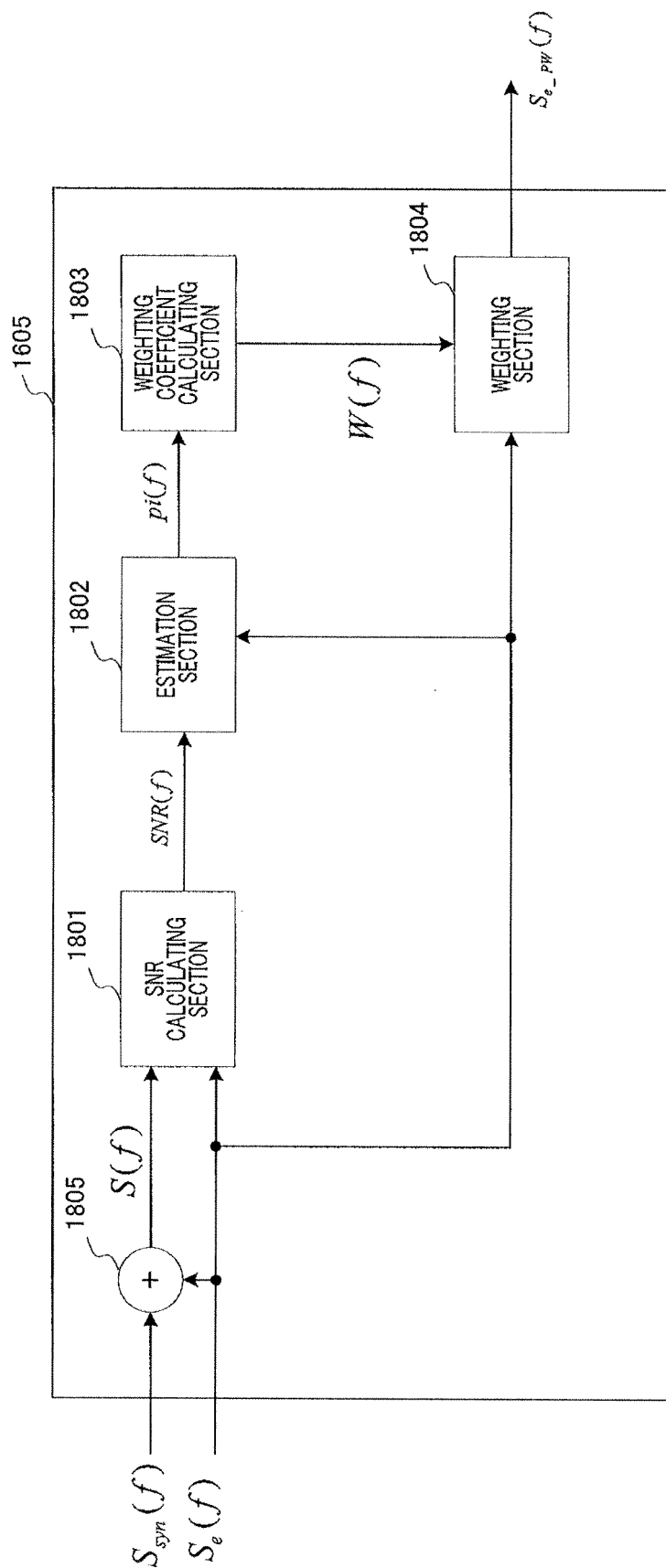


FIG.18

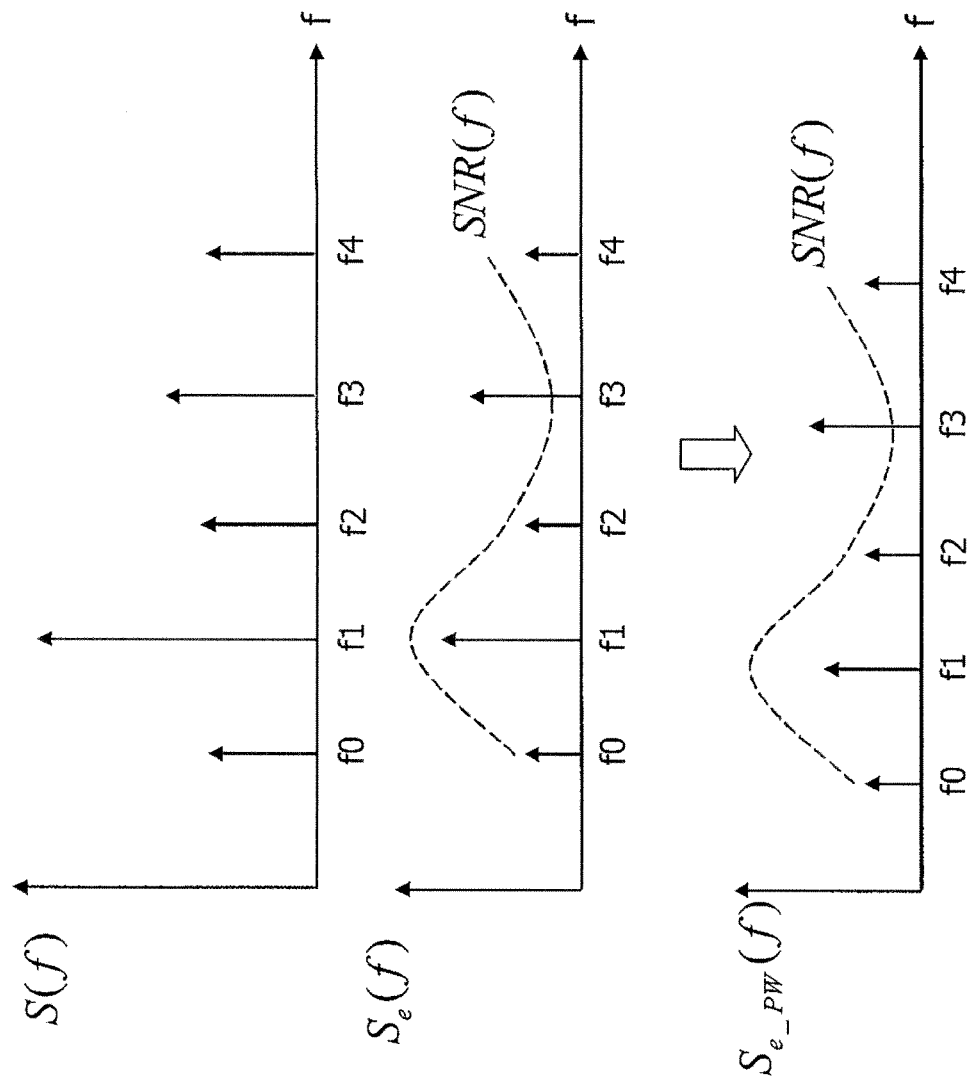


FIG.19

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 2007016404 A1 [0062]

Non-patent literature cited in the description

- **T. VAILLANCOURT et al.** ITU-T EV-VBR: A Robust 8-32 kbit/s Scalable Coder for Error Prone Telecommunication Channels. *Proc. Eusipco, Lausanne*, August 2008 [0056]
- **LEFEBVRE et al.** High quality coding of wideband audio signals using transform coded excitation (TCX). *IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1994, vol. 1, 1/193-1/196 [0056]
- **KARL HEINZ BRANDENBURG.** MP3 and AAC Explained. *AES 17th International Conference*, September 1999 [0056]